DATA DRIVEN KERNEL CHOICE IN NON-PARAMETRIC CURVE ESTIMATION

vom Fachbereich für Mathematik und Informatik der Technischen Universität Braunschweig genehmigte Dissertation zur Erlangung des Grades einer Doktorin der Naturwissenschaften (Dr. rer. nat.)

CLEMENTINE DALELANE

Präsident der Technischen Universität Braunschweig
Professor Dr. Jochen Litterst

Dekan des Fachbereichs für Mathematik und Informatik
Professor Dr. Jens-Peter Kreiss

Gutachter:

1. Professor Michael M. Neumann
2. Professor Alexandre B. Tsybakov
3. Professor Enno Mammen

eingereicht am: 26. August 2004
Tag der mündlichen Prüfung: 9. Dezember 2004

Abstract

In the case of one dimensional kernel density estimation on independent identically distributed observations, we suggest a cross-validation technique for the selection of a kernel from the class of all L_2-integrable kernel functions which have a unimodal Fourier transform. The resulting estimator is both asymptotically MISE-efficient and asymptotically minimax-adaptive over the whole scale of Sobolev classes with smoothness index greater than $1/2$, including ∞. The proofs rely on an additive decomposition of the empirical processes involved. The technique is further applied to non-parametric regression estimation. The minimax risk is established on Sobolev classes with non-integer smoothness index.

Zusammenfassung

Für den Fall der eindimensionalen Kern-Dichteschätzung bei unabhängigen, identisch verteilten Beobachtungen wird eine Cross-Validation-Technik für die Wahl des Kerns aus der Klasse aller L_2-integrierbaren Kernfunktionen mit unimodaler Fouriertransformierter vorgeschlagen. Der resultierende Schätzer ist sowohl asymptotisch MISE-effizient als auch asymptotisch minimax-adaptiv auf der gesamten Skala von Sobolev-Klassen mit Glattheitsindex größer als $1/2$, einschließlich ∞. Die Beweise stützen sich auf eine additive Zerlegung der betreffenden empirische Prozesse. Die Methode wird anschließend auf die nichtparametrische Regressionsschätzung übertragen. Das Minimax-Risiko auf Sobolev-Klassen mit nichtganzzahligen Glattheitsindex wird hergeleitet.

Résumé

Pour le cas de l'estimation à noyau d'une densité unidimensionnelle à base de données indépendantes et identiquement distribuées, on propose une technique de validation croisée pour le choix du noyau dans la classe de tout les noyaux integrables au sens L_2, dont la transformée de Fourier est unimodale. L'estimateur, qui en résulte, est asymptotiquement efficient par rapport au MISE et asymptotiquement adaptif au sens minimax pour toute l'échelle de classes de Sobolev de densités à l'indice de lissité supérieur à $1/2$, y comprit ∞. Les preuves sont basées sur une décomposition additive des processus empiriques concernés. Ensuite la méthode est appliquée à la régression nonparamétrique. Le risque minimax est déduit pour les classes de Sobolev à l'indice de lissité nonentier.

Contents

Part I
NON-PARAMETRIC DENSITY ESTIMATION

1 INTRODUCTION

This work relates to a fundamental problem in the field of non-parametric statistics: one-dimensional density estimation. The sample of observations $(X_1, X_2, \dots, X_n)$ is assumed to consist of independent, identically distributed random variables with support in $\mathbb{R}$ and density function f. A kernel density estimator $\widetilde{f}_K(x)$ is defined as:

$$\widetilde{f}_K(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(X_i - x\right) \tag{1}$$

where K is the kernel function. When K is rescaled by a bandwidth parameter $h > 0$, $K_h(.) = \frac{1}{h}K(./h)$, with $h \longrightarrow 0$ as $n \longrightarrow \infty$, but $nh \longrightarrow \infty$, then $\widetilde{f}_{K_h}$ is a consistent estimator for f. We are going to evaluate the performance of the estimator via the quadratic risk (Mean Integrated Square Error):

$$MISE(K) = E\int\left(\widetilde{f}_K(x) - f(x)\right)^2 dx \tag{2}$$

It is a familiar fact that, for a given kernel function K, only the bandwidth is relevant as regards the rate of convergence of MISE towards 0 and that kernel functions of the same order p (i.e. $\mu_p(K) = \int x^p K(x)dx < \infty$, while $\mu_{p-i}(K) = 0$ for $i = 1, \dots p-1$) yield identical convergence rates.
Cross-validation (CV) is one commonly used method to determine in practice the bandwidth. The so-called plug-in procedures provide an alternative to CV. However, as summarized in Marron, Park (1990), these methods require two additional degrees of smoothness of the unknown density function, compared to cross-validation, in order to achieve the optimal convergence rate, that is attainable by a given kernel. Moreover Hall (1993) found some bizarre, and more or less unfeasible selection rules for the bandwidth of the pilot estimate, such that in spite of all drawbacks, cross-validation and related techniques seem to be the only objective criterion of bandwidth choice.
Hall (1983) showed for the first time the efficiency of the following method, proposed in two unrelated works by Bowman (1984) and Rudemo (1982):

$$CV(h) = \int \widetilde{f}^2_{K_h}(x)dx - \frac{2}{n(n-1)}\sum_{i\neq j} K_h\left(X_i - X_j\right) \longrightarrow \min_{h>0} \tag{3}$$

In particular, the convergence in probability of the CV-optimal bandwidth, h_0, towards the optimal bandwidth in the sense of MISE minimization, h^* (i.e. $ISE(h_0)/MISE(h^*) \longrightarrow 1$ in probability), was shown for densities f with bounded second derivative and kernel functions with bounded support and $\int x^2 K''(x)dx < \infty$. Stone (1984) has proven, under weaker assumptions on f and for Hölder-continuous kernel functions with bounded support, that $ISE(h_0)/\min ISE(h) \longrightarrow 1$ with probability 1. Given that $ISE(h) \sim MISE(h)$, the difference in the definition of asymptotic optimality is irrelevant. Nolan, Pollard (1987) expanded Stone's result to kernel functions with arbitrary support, but bounded variation.

Thereafter Hall, Marron (1987) had been able to show also second order efficiency of the cross-validation procedure. Yet both approaches, cross-validation and plug-in, share the characteristic of letting the bandwidth be data-driven, while the kernel function K remains fixed.
The optimal choice of the kernel function for a given kernel order is addressed by Epanechnikov (1969) and Gasser, Müller, Mammitzsch (1985). For f in the β^{th} Sobolev space $\mathcal{S}_\beta := \{f \in L_2 | \int |f^\beta(x)|^2 dx < \infty\}$, they consider AMISE, i.e. the asymptotically leading part of the MISE $h^{2\beta}\mu_\beta^2(K)\|f^{(\beta)}\|_2^2 + (nh)^{-1}\|K\|_2^2$, and the optimal bandwidth h^* hereof (AMISE-optimal bandwidth). Since the order of the kernel function determines the optimal rate of convergence through $h^{2\beta} \sim (nh)^{-1}$, differences in AMISE between kernel functions of order β can only be reflected in the β^{th} moment and the L_2-norm of the kernels. It can be seen, from the structure of AMISE, $AMISE(h^*) = C \cdot n^{-2\beta/(2\beta+1)}\mu_\beta^2(K)\|f^{(\beta)}\|_2^2\|K\|_2^{2\beta}$, that a balance between $\mu_\beta^2(K)$ and $\|K\|_2^{2\beta}$ should be sought. Gasser, Müller, Mammitzsch (1985) therefore suggest using kernel functions that minimize this product.
Because the mentioned minimization problem has no unique solution within the class of kernel functions of order β, a pseudo-optimal kernel is chosen from a subset which contains all polynomials with $(\beta - 2)$ changes of sign, restricted to the support $[-1, 1]$, so-called minimal kernel functions (for second order the optimal minimal kernel would be the well known Epanechnikov kernel). As supposed, the optimal minimal kernel function is in no sense optimal within the class of all kernel functions of order β. And furthermore, simulation studies as in Hansen (2003) show that the substitution of AMISE for the exact finite-sample MISE can be quite misleading in many cases of interest.
Opposed to the minimal kernels, minimax kernel functions enable to control, within a given Sobolev class of density functions $\mathcal{S}_\beta(L) := \{f \in L_2 | \int |f^{(\beta)}(x)|^2 dx \le L\}$, the maximal exact MISE, see Efroimovich, Pinsker (1983), Golubev (1991) and (1992) and Schipper (1996). Though, some scepticism might persist, whether this "worst case"-approach is appropriate not only for rather pathological estimation problems, or if well behaved density functions could in general be estimated with lower MISE values.
Besides that, the minimax estimator is not meant to adapt to the possibly higher smoothness of some densities which are also contained in a given Sobolev class. The already mentioned working paper Hansen (2003) finds substantial efficiency gains (and losses) through the use of kernels of different order, so the following approach by Hall and Marron (1987) might be an appealing development.
In contrast to all contributions mentioned so far, Hall and Marron (1987) do not consider the degree of smoothness of the underlying density function as known. They construct a family of kernel functions $\{K_{h,\beta}(x)|\ h > 0,\ \beta \in \mathbb{N} \cup \infty\}$. For $\beta \in \mathbb{N}$, $K_{h,\beta}$ is defined as $K_{h,\beta}(x) = (h\pi)^{-1}\int_0^\infty \cos(tx/h)\exp\{-t^\beta\}dt$, where the first parameter denotes, as usual, the bandwidth, while the second one determines the order of the kernel function. When $\beta \longrightarrow \infty$, then $K_{h,\beta}$ converges to the sinc-kernel $(\pi x)^{-1}\sin(hx)$. The cross-validation criterion is minimized simultaneously with respect to both parameters. The CV-optimal parameters h_0, β_0 converge towards the AMISE-optimal ones, h^*, β^*. However, within given Sobolev classes of densities $\mathcal{S}_\beta(L)$, $\beta \in \mathbb{N}$, Hall and Marron (1987) have proven no optimality properties of the kernel function $K_{\cdot,\beta}$.
Having a closer look, even the choice of $\beta \in \mathbb{N}$ does not yet fully exhaust the potential revealed by the notion of Sobolev classes. Hall, Marron (1987) themselves mention a possible extension of the smoothness concept. Let $\widehat{g}$ denote the Fourier transform of a function g, if it exists (see Appendix A.1 for L_2-theory of Fourier transforms). With Parseval's equality

$$f \in \mathcal{S}_\beta(L) \quad \Longleftrightarrow \quad f \in L_2 \text{ and } \int \left(f^{(\beta)}(x)\right)^2 dx = \frac{1}{2\pi}\int |\omega^\beta \widehat{f}(\omega)|^2 d\omega \leq L$$

Instead of $\beta \in \mathbb{N}$, β could be let vary in $\mathbb{R}^+$. It is not clear, if it were meaningful also to prolong the concept of kernel order to the real positive numbers, and how this could be done. Already ten years before Hall and Marron, Davis (1977) pursued a different way of looking at the problem of kernel order. She analyzed super kernels (absolutely integrable kernels with $\widehat{K}(\omega) = 1$ in an open neighborhood of 0) and the sinc-kernel, wich are rate adaptive over scales of Sobolev classes of densities, excluding higher smoothness, the sinc-kernel being rate adaptive for all densities, even those, which do not fit to any upper smoothness bound (but all with $|\widehat{f}|$ monotonously decreasing on $\mathbb{R}^+$). That is, a kernel estimator for the density f with a super kernel or the sinc-kernel achieves the optimal rate of convergence, for f of different or all degrees of smoothness, respectively. The assumption, that $|\widehat{f}|$ decreases monotonously on $\mathbb{R}^+$, is not a very serious restriction, since this is the case with many of the common densities, as there are for example the Normal, the Cauchy, the Gamma, the Chi-square and the Exponential distribution density. And it could possibly be overcome by just minor variations in the proof of the results. But as Davis (1977) herself has also shown, super kernels do not possess the asymptotic minimax property with respect to the constants. And the sinc-kernel is asymptotically minimax exclusively for entire densities, i.e. for densities whose characteristic function has bounded support.
Unfortunately, it is a clear matter of fact, that there exists no kernel, that is uniformly MISE-optimal for all density functions, whatever skillfully the bandwidth might be chosen. In the light of this statement, the only solution appears to be a data adaptive kernel selection. So we are led back to the idea of Hall, Marron (1987), but now asking to have a much broader variety of kernel functions to our disposal.
The recent application of Stein's blockwise estimator to the problem of non-parametric density estimation by Rigollet (2004), where a linear combination of several sinc-kernels with differently fixed bandwidths is fitted to minimize an estimate of the quadratic risk, can be regarded as such a data adaptive kernel selection. The procedure is minimax-adaptive and approximates the MISE of the so-called monotone oracle, i.e. the MISE-minimizing element of a wide class of kernel functions with monotone Fourier transform on $\mathbb{R}^+$.
On closer examination we recognize that, when a kernel function is to be chosen out of a family of kernels with sufficiently general definition, it is no longer useful to distinguish between the kernel function and the bandwidth. Because all rescaled kernel functions are already included in the initial family, which is for example the case with the monotone oracle.
By Parseval's equality it is possible to determine, whenever $\|f\|_2$ exists, the Fourier transform of the MISE-optimal kernel function $\widehat{K}_{opt}$, and therewith the MISE-optimal kernel K_{opt} itself among all square integrable functions.

$$\begin{aligned} MISE(K) &= E\int_{-\infty}^{\infty}\left(\widetilde{f}_K(x) - f(x)\right)^2 dx \\ &= \int \left(E\widetilde{f}_K(x) - f(x)\right)^2 + E\left(\widetilde{f}_K(x) - E\widetilde{f}_K(x)\right)^2 dx \\ &= \int \left(K * f(x) - f(x)\right)^2 + n^{-1}\left(K^2(x) - (K*f)^2(x)\right) dx \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} |\widehat{f}(\omega)|^2 \left|\widehat{K}(\omega) - 1\right|^2 + n^{-1}|\widehat{K}(\omega)|^2\left(1 - |\widehat{f}(\omega)|^2\right) d\omega \qquad (4) \end{aligned}$$

The integrand in the last row can be minimized pointwisely. $\widehat{K}_{opt}$ is a function in L_2, which takes value 1 at point 0, its absolute value is everywhere ≤ 1 and it decreases to 0, as

$|\omega| \longrightarrow \infty$. Obviously, the optimal kernel does not only depend on the underlying density f, but also on the number n of observations. Therefore, for any density f, we have a sequence of optimal kernels, tending to the Dirac measure at point 0, when $n \longrightarrow \infty$, which has the Fourier transform $\widehat{K}(\omega) \equiv 1$. This convergence coincides with the behavior of a fixed kernel function that is rescaled by the sequence of bandwidths $h_n \longrightarrow 0$, as $n \longrightarrow \infty$. Another obvious property of any kernel K, which minimizes the MISE for some density f and some number n of observations, is that its Fourier transform is necessarily real, symmetric and non-negative.

Now let us, for a moment, consider the concept of curve smoothing from the viewpoint of signal recognition. The idea behind all kinds of smoothing techniques is to consider high frequency information of the data as caused by randomness and to replace it by 0, in order to recover low frequency patterns, that are perceived as features of the true curve. In terms of kernel density estimation, that means that the values of the empirical characteristic function of a sample at higher frequencies are suppressed by a kernel function, whose Fourier transform is 0 for high frequencies. More generally spoken, the higher the frequency, the smaller should be the data's impact on the estimator, that is allowed for. So when intending to estimate a density (or as well any other curve) by a kernel method, it is a quite reasonable assumption, that the Fourier transform of the kernel be unimodal with mode 0.

In addition, the fact that the Fourier transforms of many common densities are in absolute value themselves unimodal and therefore produce optimal kernels with unimodal Fourier transform, supports our approach by a more heuristic argument. A practical advantage of the restriction on unimodal Fourier transforms is the convexity of the resulting class of kernel functions, which will be denoted from now on by

$$\mathcal{K} := \left\{ K \in L_2 | \widehat{K} \searrow 0 \right\}, \tag{5}$$

the class of all (kernel) functions in L_2 with real, symmetric, non-negative and unimodal Fourier transform, $\widehat{K}(0) = 1$. The objective of the present work, Part I is to propose the kernel estimator $\widetilde{f}_{K_0}$ to estimate the common density of a sample $(X_1, \ldots, X_n)$ of i.i.d. observations, employing the kernel K_0, which minimizes the cross-validation criterion (instead of the unavailable MISE) over the class of kernels $\mathcal{K}$. Asymptotic MISE-efficiency among the class $\mathcal{K}$ is shown to hold whenever the unknown density function is bounded and L_2-integrable. It should be mentioned that $\widetilde{f}_{K_0}$ represents for the first time a completely data determined kernel estimator without any arbitrary restriction to the kernel, except the well justified assumption $\widehat{K} \searrow 0$. The minimax-adaptivity of the procedure over Sobolev classes with smoothness index greater than 1/2 is proven in Part III of this work, whereas a similar method of CV-kernel selection for the problem of non-parametric regression is developed in Part II.

Subsection 1.1 is dedicated to a comment on the proving techniques. But already at this stage we might venture a technical remark. It had been possible to approximate the supremum of an empirical process to which the application of chaining is not straight forward, by means of an additive decomposition. We believe it would be a very interesting question, to analyze in which cases this technique can provide an alternative to chaining.

Acknowledgement I would like to take advantage of this opportunity to express my gratitude to Prof. M. Neumann, who initiated the present work and supported it in the most valuable way. I shall further thank Prof. A. Munk for his profound discussion on general aspects of smoothing.

1.1 About the method of the proof

In order to prove the convergence in probability of the quadratic loss of the kernel estimate with CV-optimal bandwidth h_0, $ISE(h_0)$, towards the quadratic risk of the kernel estimate with MISE-optimal bandwidth h^*, $MISE(h^*)$, Hall (1983) approximates $ISE(h_0)$ by means of Gaussian processes that satisfy in turn this asymptotic condition.
Stone (1984) breaks down the difference $CV(h) - ISE(h)$ to an empirical process and an empirical degenerate U-process. The convergence of the empirical process is derived directly using Bernstein's inequality. The convergence rate of the U-process is established by means of a Poissonization argument.
A new and elegant proof of the same proposition (under slightly different assumptions) is presented by Nolan, Pollard (1987). The proposition is shown to be a special case of the developments in the theory of empirical U-processes brought about by "chaining" techniques. The characterization of the problem explicitly refrains from going beyond the one in Stone (1984).
Trying to adapt Nolan/Pollard's proof to our problem, we observe immediately that a crucial assumption to Nolan/Pollard is not fulfilled:
Consider the so-called covering number – an important magnitude that gives an insight to the size and richness of a class of functions (in our case the class of available kernel functions). The covering number $N_p(\varepsilon, Q, \mathcal{G}, \breve{g})$ (Pollard (1984)) of a class $\mathcal{G}$ of functions with envelope $\breve{g}$ with respect to a measure Q (assuming $(\int |\breve{g}|^p dQ)^{1/p} = 1$) is the cardinality of the smallest subclass $\bar{\mathcal{G}}$ of $\mathcal{G}$ satisfying

$$\forall\ g \in \mathcal{G}\ \ \exists\ \bar{g} \in \bar{\mathcal{G}} :\ \|g - \bar{g}\|_{Q,p} := \left(\int |g - \bar{g}|^p dQ \right)^{1/p} \leq \varepsilon$$

For Nolan/Pollard's argument to apply, the L_1- and the L_2-norm must have somehow bounded covering numbers:

$$N_p(\varepsilon, \mathcal{G}) = \sup_Q N_p(\varepsilon, Q, \mathcal{G}, \breve{g}) \leq A_p \varepsilon^{-V_p}, \qquad A_p \text{ and } V_p < \infty, \quad p = 1, 2$$

i.e. $\mathcal{G}$ must be a Euclidean class. This condition is satisfied for the usual type of classes of rescaled kernel functions $\mathcal{H}_K = \{K_h | h > 0\}$ with fixed K of bounded variation. But it is not difficult to see, that for the class $\mathcal{K}$, the L_1-covering numbers for $\varepsilon \leq 1/2$ are unbounded.
In Subsection 3.1, technical reason is given why it constitutes no loss of generality to replace the class $\mathcal{K}$ by the smaller class $\mathcal{K}_n := \{K \in \mathcal{K} | \operatorname{supp} \widehat{K} \in (-n, n)\}$, comprising only those elements of $\mathcal{K}$ whose Fourier transform is supported on the finite interval $(-n, n)$. This interval is obviously growing with n. And so are the L_1-covering numbers: $N_1(\varepsilon, \mathcal{K}_n) \geq \varepsilon^{-1} n$, thus not admitting the assumed polynomial bound.
Though it is certainly possible to relax Nolan/Pollard's strong constraints to the covering numbers so as to allow for bounds increasing with n, there exists also some difficulty to actually find a clearly arranged function grid $\overline{\mathcal{K}_n}$ on $\mathcal{K}_n$, that has on one hand not too a big cardinality, but on the other hand approximates the functions in $\mathcal{K}_n$ accurately enough. Ad-hoc choices, where a regular lattice of step functions is laid over the Fourier transforms $\widehat{K}$, lead to grids of a cardinality that involves terms of $O(\frac{1}{\varepsilon}!)$ (to be understood as factorial of the integer part of $\frac{1}{\varepsilon}$). But then, a bounded covering integral $\int_0^s \ln N_2(\varepsilon, \mathcal{K}_n) d\varepsilon \leq$ const. for $s \longrightarrow \infty$ is essential to chaining. In this concern, the ad-hoc grid of step functions apparently fails.
However, chaining is not necessarily the only way of approximating functions in a given class and thereby of controlling the supremum of the related empirical process. Tentatively recall

the multiresolutional analysis of wavelet transform, which is an additive decomposition on a special orthonormal function basis.
A simple and evident example demonstrates the advantages that such an additive decomposition may possess: Take a set of step functions $\mathcal{I} = \left\{\sum_{i=1}^{m} \theta_i \, I_{[i-1,i)}(x),\ \theta_i \in \{0,1\}\right\}$. To represent all step functions contained in $\mathcal{I}$ by means of a function grid, a grid of 2^m functions would be necessary. In contrast, for additive decomposition, m basis functions would suffice. Now allow for $\theta_i \in \{0,1,2\}$ for all i. Obviously, the additive decomposition basis would serve as well as before. Whereas the function grid would have to be enlarged up to more than twice as many grid functions, namely 3^m.
Hereof we learn that in terms of additive decomposition the richness of a function class is apportioned between the richness of the set of basis functions the richness of the set of sequences of values which are taken by the decomposition coefficients.
We have been able to benefit from this property in the proof of our main theorem (efficiency of kernel CV). Its central idea consists of decomposing the empirical process and the empirical U-process, which were already distinguished by Stone (1984), into linear combinations of a countable basis of empirical processes and empirical U-processes, respectively (Subsections 4.2 and 4.1). Separate arguments as regards a bound for the empirical basis processes, not depending on K, and a bound for the decomposition coefficients, which are non-random, are combined so as to result in the desired convergence.
By means of Bernstein's inequality and a Bernstein type inequality that applies to empirical degenerate U-processes, shown by Arcones and Giné (1993), the probability of the subset of samples $(X_1, \ldots, X_n)$, where the absolute value of at least one basis process exceeds a given threshold, is shown to be bounded by $n^{-\lambda}$, where λ is an arbitrarily large constant (Lemmata I.4.2 and I.4.4). In turn, the sums of absolute coefficients are small due to the monotony of the Fourier transforms of the $K \in \mathbb{K}$ (Lemmata I.4.1 and I.4.3).
In Section 2 the exact hypotheses and the result are presented. Section 3 contains the proof of the theorem stated in Section 2, relying on 3 propositions: **A1**, **A2** and **A3**. In Sections 4 and 5, propositions **A1**, **A2** and **A3** of Section 3 are proven. Finally, Section 6 considers possibilities for the practical computation of the optimal kernel function.

2 MAIN THEOREM

Let $(X_1, X_2, \ldots, X_n) \in \mathbb{R}^n$ be an i.i.d. sample with common density function f. Let $\widetilde{f}_K(x)$, $MISE(K)$ and $CV(K)$ be defined as in Section 1, definitions (1), (2) and (3).
Let the density f be bounded, $\|f\|_\infty < \infty$, have finite L_2-norm, $\|f\|_2 < \infty$, and denote by $\widehat{f}(\omega) = \int f(x) e^{ix\omega} dx$ the characteristic function of f.
Let $\mathcal{K}$ be (again as in Section 1, definition (5)) the set of all kernel functions with real, symmetric, non-negative and unimodal Fourier transform $\{K \in L_2 | \widehat{K} \searrow 0\}$. Furthermore, for technical reason let $\|K\|_2$ be $\leq \sqrt{n}$. Let K^* be the MISE-optimal kernel function for f and n among the class $\mathcal{K}$, and let K_0 be the CV-optimal kernel function among $\mathcal{K}$ restricted to kernels, whose Fourier transform additionally has support in $(-n, n)$.

$$\begin{aligned} K^* &:= \arg\min \left\{ MISE(K) \middle| K \in \mathcal{K} \right\} \\ K_0 &:= \arg\min \left\{ CV(K) \middle| K \in \mathcal{K},\ \operatorname{supp} \widehat{K} \subseteq (-n, n) \right\} \end{aligned} \tag{6}$$

The technical remark of Subsection 3.1 will explain, why this assumption is irrelevant to the generality of the statement.

THEOREM I.2.1 Under these hypotheses, for all $\delta > 0$ it holds that:

$$|E[ISE(K_0)] - MISE(K^*)| = O(n^{-\delta})MISE(K^*) + O\left(n^{\delta-1}\right) \tag{7}$$

Remark The residuals $O(n^{-\delta})$ and $O(n^{\delta-1})$ in Theorem I.2.1 contain two constants depending on f: $\|f\|_2$ and $\max f$. Since obviously $\int f^2(x)dx = \frac{1}{2\pi}\int|\widehat{f}(\omega)|^2 d\omega \leq \frac{1}{2\pi}\int|\widehat{f}(\omega)|d\omega$, $\max|f| \leq \frac{1}{2\pi}\int|\widehat{f}(\omega)|d\omega$ and furthermore $\int|\widehat{f}(\omega)|d\omega \leq 2+\varepsilon^{-1/2}(\int|\omega^{1/2+\varepsilon}\widehat{f}(\omega)|^2 d\omega)^{1/2}$, these are uniformly bounded over Sobolev classes $\mathcal{S}_\beta(L)$, i.e. $\frac{1}{2\pi}\int|\omega^\beta\widehat{f}(\omega)|^2 d\omega \leq L$, with smoothness index $\beta > 1/2$.

3 PROOF

We assume the following propositions, which will be proven in Sections 4 and 5: For any $\lambda < \infty$, there exists a set $A_n \subseteq \mathbb{R}^n$, such that for an arbitrary observation $X = (X_1, \ldots, X_n) \in A_n$ and $\delta > 0$ it holds that:

A1 $\quad |ISE(K) - \widetilde{CV}(K)| = O(n^{-\delta})MISE(K) + O\left(n^{\delta-1}\right) \qquad \forall K \in \mathcal{K},\ \operatorname{supp}\widehat{K} \subseteq (-n,n)$

A2 $\quad |ISE(K) - MISE(K)| = O(n^{-\delta})MISE(K) + O\left(n^{\delta-1}\right) \quad \forall K \in \mathcal{K},\ \operatorname{supp}\widehat{K} \subseteq (-n,n)$

A3 $\quad P\left(X \in A_n^c\right) = O\left(n^{-\lambda}\right)$

where $O(n^{-\delta})$ and $O(n^{\delta-1})$ do not depend on K. In case f is a density function that can only be estimated at a rate $n^{\varepsilon-1}$, $n^{-\delta}MISE(K)$ will dominate $n^{\delta-1}$ for small enough $\delta > 0$ at the right-hand side of these equations. Otherwise, either if δ is too big or if f can be estimated at a rate of $n^{-1}\ln n$ (for example Normal and Cauchy distribution density) or even at a rate of n^{-1} (entire density), the term $n^{\delta-1}$ will dominate the others.
$\widetilde{CV}$ is a criterion derived from CV, such that $\widetilde{CV}(K) - \widetilde{CV}(K') = CV(K) - CV(K')$ for any K, K' in $\mathcal{K}$, and will be defined below. In addition, at the end of this section it will be proven that: $ISE(K) \leq (\|K\|_2 + \|f\|_2)^2 \leq (n^{1/2} + \|f\|_2)^2$. For a set $C \subseteq \mathbb{R}^n$ and a function g defined on $\mathbb{R}^n$, let $E_C[g(X)]$ denote the expectation $E[g(X)\cdot I(X \in C)]$. As a consequence, we can proceed in the following way:

$$\begin{aligned}
& E[ISE(K_0)] - MISE(K^*) \\
= \;& E\Big[ISE(K_0) - ISE(K^*)\Big] \\
\leq \;& E_{A_n}\Big[ISE(K_0) - ISE(K^*)\Big] + P\left(A_n^c\right)\sup_{K\in\mathbb{K}} ISE(K) \\
= \;& E_{A_n}\Big[ISE(K_0) - \widetilde{CV}(K_0) + \widetilde{CV}(K_0) - \widetilde{CV}(K^*) + \widetilde{CV}(K^*) - ISE(K^*)\Big] \\
& + O\left(n^{-\lambda}\right)\left(\sqrt{n} + \|f\|_2\right)^2 && (\mathbf{A3}) \\
\leq \;& E_{A_n}\Big[ISE(K_0) - \widetilde{CV}(K_0)\Big] + 0 + E_{A_n}\Big[\widetilde{CV}(K^*) - ISE(K^*)\Big] + O\left(n^{-\lambda+1}\right) \\
= \;& O(n^{-\delta})E_{A_n}[MISE(K_0)] + O(n^{-\delta})MISE(K^*) + O\left(n^{\delta-1}\right) + O\left(n^{-\lambda+1}\right) && (\mathbf{A1}) \\
= \;& O(n^{-\delta})E_{A_n}[ISE(K_0)] + O(n^{-\delta})MISE(K^*) + O\left(n^{\delta-1}\right) && (\mathbf{A2})
\end{aligned}$$

for λ sufficiently large. In order to return to $E[ISE(K_0)]$, we exert again proposition **A3** so as to find $|E_{A_n}[ISE(K_0)] - E[ISE(K_0)]| = O(n^{-\lambda+1})$, and therewith

$$\begin{aligned} E[ISE(K_0)] - MISE(K^*) &= O(n^{-\delta})E[ISE(K_0)] + O(n^{-\delta})MISE(K^*) + O\Big(n^{\delta-1}\Big) \\ \Longrightarrow \qquad E[ISE(K_0)] &= \Big(1 + O(n^{-\delta})\Big) MISE(K^*) + O\Big(n^{\delta-1}\Big) \end{aligned}$$

By **A2**, the opposite is also readily shown:

$$\begin{aligned} MISE(K^*) - E[ISE(K_0)] &= O(n^{-\delta})E[ISE(K_0)] + O\Big(n^{\delta-1}\Big) \\ \Longrightarrow \qquad MISE(K^*) &= \Big(1 + O(n^{-\delta})\Big) E[ISE(K_0)] + O\Big(n^{\delta-1}\Big) \end{aligned}$$

which implies the desired result:

$$|E\left[ISE(K_0)\right] - MISE(K^*)| = O(n^{-\delta})MISE(K^*) + O\Big(n^{\delta-1}\Big) \tag{8}$$

To complete the proof, it should be noted that:

$$\begin{aligned} ISE(K) &= \int \widetilde{f}_K^2(x)dx - 2\int \widetilde{f}_K(x)f(x)dx + \int f^2(x)dx \\ &= \int \frac{1}{n^2}\sum_{i,j} K(X_i - x)K(X_j - x)dx - \int \frac{2}{n}\sum_{i=1}^{n} K(X_i - x)f(x)dx + \int f^2(x)dx \\ &\le \int K^2(x)dx + 2\sqrt{\int K^2(x)dx}\sqrt{\int f^2(x)dx} + \int f^2(x)dx \\ &= \left(\sqrt{\int K^2(x)dx} + \sqrt{\int f^2(x)dx}\right)^2 \\ &\le \left(\sqrt{n} + \sqrt{\int f^2(x)dx}\right)^2 \end{aligned}$$

□

3.1 A technical remark

Instead of considering all kernel functions K of $\{K|\widehat{K} \searrow 0\}$, it is equivalent to withdraw the attention to a subset of $\mathcal{K}$, $\mathcal{K}_n := \{K \in \mathcal{K}|\operatorname{supp}\widehat{K} \subseteq (-n, n)\}$. Let $K^*_{(n)}$ be the truncated version of the MISE-optimal kernel K^*:

$$K^*_{(n)}(x) = \frac{1}{2\pi}\int \widehat{K}^*_{(n)}(\omega)e^{-ix\omega}d\omega\ , \quad \widehat{K}^*_{(n)}(\omega) \quad := \quad \begin{cases} \widehat{K}^*(\omega), & |\omega| < n \\ 0, & \text{otherwise} \end{cases}$$

It will be seen that under the conditions of Theorem I.2.1, namely $\|f\|_2 < \infty$, $MISE(K^*_{(n)}) = MISE(K^*)(1 + o(1))$. To limit the minimization of MISE onto $\mathcal{K}_n$ signifies therefore no loss of generality.

Proof: Transforming MISE into the Fourier domain, we can quantify the consequences of trimming the support of $\widehat{K}$ in a plain formula. From equality (4) we know:

$$MISE(K) \quad = \quad \frac{1}{2\pi}\int_{-\infty}^{\infty} |\widehat{f}(\omega)|^2 \left(\widehat{K}(\omega) - 1\right)^2 + \frac{1}{n}\widehat{K}^2(\omega)\left(1 - |\widehat{f}(\omega)|^2\right) d\omega$$

which gives in combination with the symmetry of $\widehat{K}^*$, its monotony on $\mathbb{R}^+$ and $0 \le \widehat{K}^2(\omega) \le 1$:

$$\begin{aligned}
MISE(K^*) &= \frac{1}{\pi}\int_0^\infty |\widehat{f}(\omega)|^2 \left(\widehat{K}^*(\omega)-1\right)^2 + \frac{1}{n}\widehat{K}^*(\omega)^2\left(1-|\widehat{f}(\omega)|^2\right) d\omega \\
&\ge \frac{1}{\pi}\left[\int_n^\infty |\widehat{f}(\omega)|^2 \left(\widehat{K}^*(\omega)-1\right)^2 d\omega + \int_0^n \frac{1}{n}\widehat{K}^*(\omega)^2\left(1-|\widehat{f}(\omega)|^2\right) d\omega\right] \\
&\ge \frac{1}{\pi}\left[\left(\widehat{K}^*(n)-1\right)^2 \int_n^\infty |\widehat{f}(\omega)|^2 d\omega + \widehat{K}^*(n)^2\left(1-\frac{1}{n}\int_0^n |\widehat{f}(\omega)|^2\right) d\omega\right] \\
&\ge \frac{1}{\pi}\left[\left(\widehat{K}^*(n)-1\right)^2 \int_n^\infty |\widehat{f}(\omega)|^2 d\omega + \widehat{K}^*(n)^2\left(1-\frac{1}{n}\int_0^\infty |\widehat{f}(\omega)|^2\right)_+ d\omega\right] \qquad (9)
\end{aligned}$$

For $f \in \mathcal{S}_\beta(L)$, $\beta > 1/2$, there exists $n_0 < \infty$, so that for all $n \ge n_0$ the factor $(1 - \frac{1}{n}\int_0^\infty |\widehat{f}(\omega)|^2)_+$ is positive and admits the representation $1 + o(1)$. Since both summands in (9) are positive, we can for $n \ge n_0$ write on one hand:

$$\widehat{K}^*(n) \le \sqrt{\pi(1+o(1))MISE(K^*)}$$

This $o(1)$ is uniform for $f \in \mathcal{S}_\beta(L)$, $\beta > 1/2$. On the other hand:

$$MISE(K^*) \ge \left(\widehat{K}^*(n)-1\right)^2 \frac{1}{\pi}\int_n^\infty |\widehat{f}(\omega)|^2 d\omega$$

Considering the difference between MISE of the optimal kernel and MISE of its truncated version in proportion to $MISE(K^*)$, we deduce:

$$\begin{aligned}
0 &< \frac{MISE(K^*_{(n)}) - MISE(K^*)}{MISE(K^*)} \\
&= \frac{\frac{1}{\pi}\int_n^\infty |\widehat{f}(\omega)|^2\left(2\widehat{K}^*(\omega)-\widehat{K}^*(\omega)^2\right) - \frac{1}{n}\widehat{K}^*(\omega)^2\left(1-|\widehat{f}(\omega)|^2\right) d\omega}{MISE(K^*)} \\
&\le \frac{\frac{1}{\pi}\int_n^\infty |\widehat{f}(\omega)|^2\ 2\widehat{K}^*(\omega)d\omega}{MISE(K^*)} \\
&\le \frac{2\widehat{K}^*(n)\frac{1}{\pi}\int_n^\infty |\widehat{f}(\omega)|^2 d\omega}{MISE(K^*)} \\
&\le \frac{2\widehat{K}^*(n)\frac{1}{\pi}\int_n^\infty |\widehat{f}(\omega)|^2 d\omega}{\left(\widehat{K}^*(n)-1\right)^2\frac{1}{\pi}\int_n^\infty |\widehat{f}(\omega)|^2 d\omega} \\
&= \frac{2\sqrt{\pi(1+o(1))MISE(K^*)}}{\left(1-\sqrt{\pi(1+o(1))MISE(K^*)}\right)^2} \\
&= \left(2\sqrt{\pi}+o(1)\right)\sqrt{MISE(K^*)}
\end{aligned}$$

which decreases to 0 whenever $MISE(K^*)$ is a decreasing sequence. Moreover, the decay of $o(1)$ is uniform, when $f \in \mathcal{S}_\beta(L)$, $\beta > 1/2$. □

The decision to restrict the support of the Fourier transforms of the kernel functions to the interval $(-n, n)$ is but one possibility, chosen for convenience. Any interval of polynomial rate $(-n^\gamma, n^\gamma)$, $\gamma > 1/2$, would serve our purpose to fulfill $MISE(K^*_{(n^\gamma)}) - MISE(K^*) =$

$o(1)MISE(K^*)$. For $\gamma \in (1/2, 1)$, the same calculation as above would yield the result. For $\gamma > 1$, evidently $MISE(K^*_{(n^\gamma)}) - MISE(K^*) < MISE(K^*_{(n)}) - MISE(K^*)$. Nevertheless the polynomial rate cannot be replaced by an exponential one, which can be seen from the proofs of Sections 4 and 5.

4 ISE-CV

From this point through the whole proof of propositions **A1**, **A2** and **A3**, the notion $O(.)$ will exclusively be used for terms not depending on f, so that the uniformity of the results over Sobolev classes of densities will be easy to retrace.

Let $X_1, \ldots, X_n$ be distributed as assumed in Section 2. Let X and Y denote two further random variables with the same distribution, independent of $X_1, \ldots, X_n$ and of each other.

$$\begin{aligned}
ISE(K) &:= \int \left(\widetilde{f}_K(x) - f(x)\right)^2 dx \\
&= \int \widetilde{f}_K^2(x)dx - 2\int \widetilde{f}_K(x)f(x)dx + \int f^2(x)dx \\
&= \int \widetilde{f}_K^2(x)dx - \frac{2}{n}\sum_{i=1}^n \int K(X_i - x)f(x)dx + \int f^2(x)dx \\
&= \int \widetilde{f}_K^2(x)dx - \frac{2}{n}\sum_{i=1}^n E\left[K(X_i - X)|X_i\right] + E\left[f(X)\right] \qquad (10) \\
CV(K) &= \int \widetilde{f}_K^2(x)dx - \frac{2}{n(n-1)}\sum_{i\neq j} K(X_i - X_j)
\end{aligned}$$

We obtain $\widetilde{CV}$ from CV by adding a zero and a further term which does not depend on K. Define $I_n(\omega) := I(|\omega| < n)$ and

$$h_f(x) := \frac{1}{2\pi}\int \widehat{f}(\omega)\left(1 - I_n(\omega)\right)e^{-i\omega x}d\omega \qquad (11)$$

the high-frequency contribution of $\widehat{f}$ to f.

$$\begin{aligned}
\widetilde{CV}(K) := CV(K) &+ \Bigg[\frac{2}{n}\sum_{j=1}^n E\left[K(X - X_j)|X_j\right] - 2E\left[K(X - Y)\right] \\
&- \frac{2}{n}\sum_{j_1}^n E\left[K(X - X_j)|X_j\right] + 2E\left[K(X - Y)\right]\Bigg] + \frac{2}{n}\sum_{j=1}^n \left(f(X_j) - h_f(X_j)\right) \\
&- 2E\Big[f(X_j) - h_f(X_j)\Big] \qquad (12)
\end{aligned}$$

We can now split the difference between the quadratic loss and the cross-validation criterion into two summands:

$$\begin{aligned}
&ISE(K) - \widetilde{CV}(K) \\
= &-\frac{2}{n}\sum_{i=1}^n E\left[K(X_i - Y)|X_i\right] + E\left[f(X)\right] + \frac{2}{n(n-1)}\sum_{i\neq j} K(X_i - X_j)
\end{aligned}$$

$$
\begin{aligned}
& - \frac{2}{n}\sum_{j=1}^{n} E\left[K(X-X_j)|X_j\right] + 2E\left[K(X-Y)\right] + \frac{2}{n}\sum_{j=1}^{n} E\left[K(X-X_j)|X_j\right] \\
& - 2E\left[K(X-Y)\right] - \frac{2}{n}\sum_{j=1}^{n}\Big(f(X_j)-h_f(X_j)\Big) + 2E\Big[f(X_j)-h_f(X_j)\Big] \\
= \quad & \frac{2}{n(n-1)}\sum_{i\neq j}\Big(K(X_i-X_j) - E\left[K(X_i-X_j)|X_i\right] - E\left[K(X_i-X_j)|X_j\right] + E\left[K(X_i-X_j)\right]\Big) \\
& + \frac{2}{n}\sum_{j=1}^{n}\Big(E\left[K(X-X_j)|X_j\right] - f(X_j) + h_f(X_j)\Big) \\
& - 2\Big(E\left[K(X-X_j)\right] - E\Big[f(X_j)-h_f(X_j)\Big]\Big) \\
=: \quad & \frac{2}{n(n-1)}\sum_{i\neq j} U_K(X_i,X_j) + \frac{2}{n}\sum_{j=1}^{n}\Big(b_K(X_j)+h_f(X_j) - E\Big[b_K(X_j)+h_f(X_j)\Big]\Big) \qquad (13)
\end{aligned}
$$

The first term corresponds to a degenerate U-statistic, since $E[U_K(X,Y)|Y] = E[U_K(X,Y)|X] = E[U_K(X,Y)] = 0$ for all values of X and Y. In Subsection 4.1 we will prove, through a wavelet decomposition, that the following holds on a certain set of "favorable events" A_{n1} in the space R^n:

$$
\frac{1}{n(n-1)}\sum_{i\neq j} U_K(X_i,X_j) = O\left(\frac{\ln^{5/2} n}{n}\right)\left(\sqrt{\int K^2(x)dx} + 1\right) \qquad (14)
$$

and for $\lambda < \infty$

$$
P\left(A_{n1}^c\right) = O\left(n^{-\lambda^{2/3}+1}\right)
$$

On the other hand, we obtain in (13) the empirical process of $b_K + h_f$ (the component of b_K that really depends on K, see equation (22)), which will be bounded on another set of "favorable events" $A_{n2} \subseteq \mathbb{R}^n$ in Subsection 4.2.

$$
\frac{1}{n}\sum_{j=1}^{n}\Big(b_K(X_j)+h_f(X_j)\Big) - E\Big[b_K(X_j)+h_f(X_j)\Big]\Big) = O\left(\frac{\ln^2 n}{\sqrt{n}}\right)\left(\sqrt{\int b_K^2(x)dx} + \frac{\|f\|_2}{\sqrt{n}}\right) \qquad (15)
$$

and

$$
P\left(A_{n2}^c\right) = O\left(n^{-\lambda+1}\right)
$$

The intersection of these two sets of "favorable events" yields the set $A_n := A_{n1} \cap A_{n2}$, used in Section 3 to bound $\widetilde{\text{CV}}$-ISE (on the very same subset A_n, ISE-MISE will be bounded in Section 5).

The bound for the U-statistic is of order $n^{-1/2}\ln^{5/2} n\sqrt{MISE(K)}$ and the one for the bias of order $n^{-1/2}\ln^2 n\sqrt{MISE(K)}$. When f is not too smooth, we find a suitable $\delta > 0$, such that this is inferior to $O(n^{-\delta})MISE(K)$. Hence:

$$
\begin{aligned}
& |ISE(K) - \widetilde{CV}(K)| \\
\leq \quad & 2\Big|\frac{1}{n(n-1)}\sum_{i\neq j} U_K(X_i,X_j)\Big| + 2\Big|\frac{1}{n}\sum_{j=1}^{n}\Big(b_K(X_j)+h_f(X_j)\Big) - E\Big[b_K(X_j)+h_f(X_j)\Big]\Big| \\
= \quad & O\left(\frac{\ln^{5/2} n}{n}\right)\left(\sqrt{\int K^2(x)dx} + 1\right) + O\left(\frac{\ln^2 n}{\sqrt{n}}\right)\left(\sqrt{\int b_K^2(x)dx} + \frac{\|f\|_2}{\sqrt{n}}\right)
\end{aligned}
$$

$$= \quad O(n^{-\delta})MISE(K) + O\left(\frac{\ln^{5/2} n \; (1+\|f\|_2)}{n}\right) \qquad \Box \; \mathbf{A1}$$

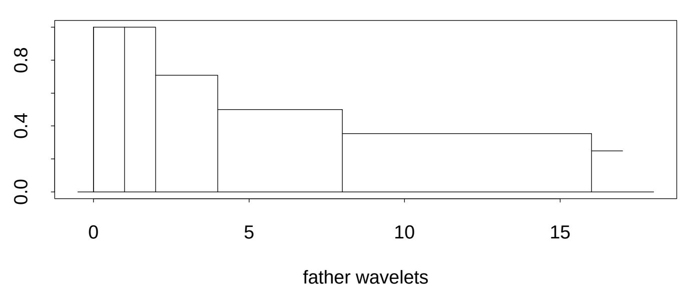
and

$$P\left(A_n^c\right) \quad \leq \quad P\left(A_{n1}^c\right) + P\left(A_{n2}^c\right) \;=\; O\left(n^{-\lambda'}\right) \qquad \text{for an appropriate } \lambda' < \infty \qquad \Box \; \mathbf{A3}$$

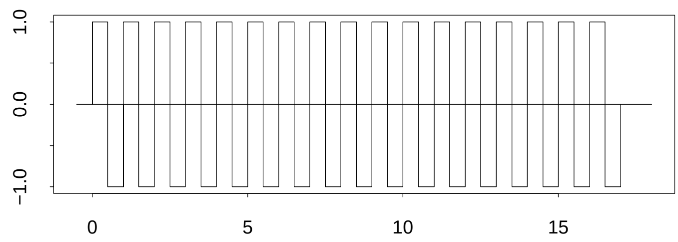
4.1 The empirical U-process

The aim of this section is to find a set A_{n1} of events where the U-statistic for every available kernel function remains "small", but whose complement shows a rapidly decreasing probability in n. This is achieved through a decomposition of U_K into a linear combination of U-processes of basis functions. We make use of the possibility of finding a bound to the probability that the U-statistic of a single basis function exceeds a certain value, by means of an inequality of Bernstein type. Through an appropriate choice of the bound, we will be able to control the probability of the event that at least one of the basis U-statistics exceeds this bound.

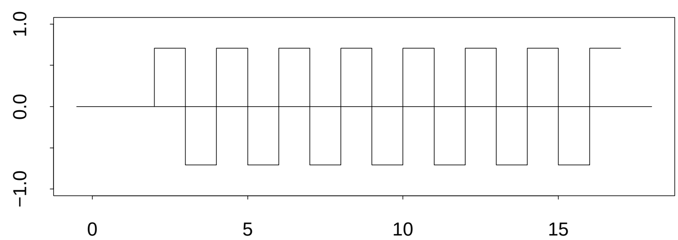
As the class $\mathcal{K}$ was defined in the Fourier domain, we know the properties of the Fourier transforms of the kernels in $\mathcal{K}$ much better than the properties of the kernel functions themselves in the time domain. We thus construct the desired basis in the Fourier domain. Inspired by the well known Haar basis, father wavelets are defined in the usual way on the interval $(-n, n)$: $\widehat{\varphi}_{0t}(\omega) := 2^{-1/2} I(|\omega| \in [t-1, t))$, $t = 1, \ldots, n$. Since the mere number of the so defined father wavelets, n, would cause problems in the proofs to come, we replace the sequence of father wavelets $(\widehat{\varphi}_{0,2^s+1}, \ldots, \widehat{\varphi}_{0,2^{s+1}})$ by just one father wavelet $\widehat{\varphi}_{-s,2}$, where $\widehat{\varphi}_{-s,2}(\omega) := 2^{-(s+1)/2} I(|\omega| \in [2^s, 2^{s+1}))$. We execute this operation for every $1 \leq s \leq d_n$, where d_n is equal to $\ln n/\ln 2$, if $\ln n/\ln 2 \in \mathbb{N}$, otherwise, $d_n := \lceil \ln n/\ln 2 \rceil$, the smallest integer greater than $\ln n/\ln 2$.



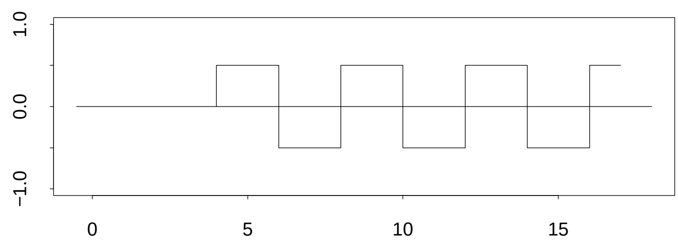
We thus result in a sequence of father wavelets $(\widehat{\varphi}_{01}, \widehat{\varphi}_{02}, \widehat{\varphi}_{12}, \widehat{\varphi}_{22}, \ldots, \widehat{\varphi}_{d_n,2})$ of $d_n + 2$ elements. On the supporting interval of every father wavelet, the mother wavelets are defined on refining scales. With notation $I_{ut}(\omega) := I(|\omega| \in [2^{-u}(t-1), 2^{-u}t))$, the mother wavelets on $(-2^{s+1}, -2^s] \cup [2^s, 2^{s+1})$ are $\widehat{\psi}_{u,t}(\omega) := 2^{(u-1)/2}[I_{u+1,2t-1}(\omega) - I_{u+1,2t}(\omega)]$, $u = -s, -s+1, \ldots, 0, 1, 2, \ldots$ and $t = 2^{s+u}+1, \ldots, 2^{s+u+1}$. When we combine all mother wavelets with the same scale index, we arrive at a sequence of $(\widehat{\psi}_{s,2}, \ldots, \widehat{\psi}_{s,2^s n})$ for $s = -1, \ldots, -d_n$, and $(\widehat{\psi}_{s,1}, \ldots, \widehat{\psi}_{s,2^s n})$, for $s \geq 0$. (Whenever there occurs an index $2^s n \notin \mathbb{N}$, it is replaced by the successing integer.) We observe that for $s < 0$, the corresponding mother wavelets do not cover the whole interval $(-n, n)$, but only $(-n, -2^s] \cup [2^s, n)$.



mother wavelets, scale 0



mother wavelets, scale −1



mother wavelets, scale −2

Unifying the notation:

$$\begin{aligned} I_{st}(\omega) &:= I\left(|\omega| \in \left[2^{-s}(t-1), 2^{-s}t\right)\right) \\ \widehat{\varphi}_{st}(\omega) &:= 2^{(s-1)/2} I_{st}(\omega) \\ \widehat{\psi}_{st}(\omega) &:= 2^{(s-1)/2}\left[I_{s+1,2t-1}(\omega) - I_{s+1,2t}(\omega)\right], \end{aligned} \tag{16}$$

we have the following complete orthonomal function basis of $L_2([-n,n])$: $\{\widehat{\varphi}_{01}\} \cup \{\widehat{\varphi}_{s2}|s = 0,\ldots,d_n\} \cup \{\widehat{\psi}_{st}|s = -1,\ldots,-d_n \text{ and } t = 2,\ldots 2^s n\} \cup \{\widehat{\psi}_{st}|s \geq 0 \text{ and } t = 1,\ldots,2^s n\}$. The decomposition of $\widehat{K}$ results in:

$$\begin{aligned} \widehat{K}(\omega) &= \alpha_{01}(K)\widehat{\varphi}_{01}(\omega) + \sum_{s=0}^{-d_n} \alpha_{s2}(K)\widehat{\varphi}_{s2}(\omega) + \sum_{s=-1}^{-d_n}\sum_{t=2}^{2^s n} \beta_{st}(K)\widehat{\psi}_{st}(\omega) + \sum_{s=0}^{\infty}\sum_{t=1}^{2^s n} \beta_{st}(K)\widehat{\psi}_{st}(\omega) \\ \alpha_{st}(K) &:= \int \widehat{\varphi}_{st}(\omega)\widehat{K}(\omega)d\omega \qquad \text{and} \qquad \beta_{st}(K) := \int \widehat{\psi}_{st}(\omega)\widehat{K}(\omega)d\omega \end{aligned} \tag{17}$$

By an inverse Fourier transform, the additive decomposition of $\widehat{K}$ can be transformed to the time domain.

$$K(x) = \alpha_{01}(K)\varphi_{01}(x) + \sum_{s=0}^{-d_n} \alpha_{s2}(K)\varphi_{s2}(x) + \sum_{s=-1}^{-d_n}\sum_{t=2}^{2^s n} \beta_{st}(K)\psi_{st}(x) + \sum_{s=0}^{\infty}\sum_{t=1}^{2^s n} \beta_{st}(K)\psi_{st}(x)$$

Accordingly, the summands in the U-process decompose into:

$$
\begin{aligned}
U_K(X_i,X_j) &= \alpha_{01}(K)U_{\varphi_{01}}(X_i,X_j) + \sum_{s=0}^{-d_n}\alpha_{s2}(K)U_{\varphi_{s2}}(X_i,X_j) + \sum_{s=-1}^{-d_n}\sum_{t=2}^{2^s n}\beta_{st}(K)U_{\psi_{st}}(X_i,X_j) \\
&\quad + \sum_{s=0}^{\infty}\sum_{t=1}^{2^s n}\beta_{st}(K)U_{\psi_{st}}(X_i,X_j),
\end{aligned}
$$

where $U_{\varphi_{st}}(X_i,X_j) := \varphi_{st}(X_i - X_j) - E[\varphi_{st}(X_i - X_j)|X_i] - E[\varphi_{st}(X_i - X_j)|X_j] + E[\varphi_{st}(X_i - X_j)]$, and $U_{\psi_{st}}$ equally defined for ψ_{st}. Interchanging the order of summation, we obtain that:

$$
\begin{aligned}
\frac{1}{n(n-1)}\sum_{i\neq j}U_K(X_i,X_j) &= \alpha_{01}(K)\left[\frac{1}{n(n-1)}\sum_{i\neq j}U_{\varphi_{01}}(X_i,X_j)\right] \\
&\quad + \sum_{s=0}^{-d_n}\alpha_{s2}(K)\left[\frac{1}{n(n-1)}\sum_{i\neq j}U_{\varphi_{s2}}(X_i,X_j)\right] \\
&\quad + \sum_{s=-1}^{-d_n}\sum_{t=2}^{2^s n}\beta_{st}(K)\left[\frac{1}{n(n-1)}\sum_{i\neq j}U_{\psi_{st}}(X_i,X_j)\right] \\
&\quad + \sum_{s=0}^{\infty}\sum_{t=1}^{2^s n}\beta_{st}(K)\left[\frac{1}{n(n-1)}\sum_{i\neq j}U_{\psi_{st}}(X_i,X_j)\right] \qquad (18)
\end{aligned}
$$

From this point onwards, the sums of wavelet coefficients and the U-statistics can be handled separately. The α's and β's are deterministic and we show in Lemma I.4.1:

$$
\begin{aligned}
|\alpha_{01}(K)| + \sum_{s=0}^{-d_n}|\alpha_{s2}(K)| &\leq \sqrt{d_n+2}\,\sqrt{\frac{1}{2\pi}\int K^2(x)dx} \\
\sum_{t=2}^{2^s n}|\beta_{st}(K)| &\leq \sqrt{\frac{1}{2\pi}\int K^2(x)dx} \qquad \text{for } s<0 \\
\sum_{t=1}^{2^s n}|\beta_{st}(K)| &\leq 2^{(-s+1)/2} \qquad \text{for } s \geq 0
\end{aligned}
$$

For a suitable constant $\lambda < \infty$, we choose our set of "favorable events" as:

$$
\begin{aligned}
A_{n1} := \Big\{(X_1,\ldots,X_n) : \quad & \left|\tfrac{1}{n(n-1)}\sum_{i\neq j}U_{\varphi_{st}}(X_i,X_j)\right| \leq \tfrac{\lambda \ln^{3/2} n}{n}, \quad (s,t) = (-d_n,2),\ldots,(0,2),(0,1); \\
& \left|\tfrac{1}{n(n-1)}\sum_{i\neq j}U_{\psi_{st}}(X_i,X_j)\right| \leq \tfrac{\lambda \ln^{3/2} n}{n}, \quad s = -d_n,\ldots,-1,\ t = 2,\ldots,2^s n; \\
& \left|\tfrac{1}{n(n-1)}\sum_{i\neq j}U_{\psi_{st}}(X_i,X_j)\right| \leq \tfrac{\lambda \ln n + s}{n}, \quad s \geq 0,\ t = 1,\ldots,2^s n\Big\} \qquad (19)
\end{aligned}
$$

whereupon the U-statistics do not become excessively large. The fact that the complement of the set A_{n1} has probability tending to 0, as $n \longrightarrow \infty$, $P(A_{n1}^c) = O(n^{-\lambda^{2/3}+1})$ uniformly for $f \in \mathcal{S}_\beta(L)$ with $\beta > 1/2$, will be shown in Lemma I.4.2 and inequality (21).

On A_{n1} it holds that (in connection with (18)):

$$\begin{aligned}
\Big|\frac{1}{n(n-1)}\sum_{i\neq j}U_K(X_i,X_j)\Big| &\leq \frac{\lambda\ln^{3/2}n}{n}\left[|\alpha_{01}(K)|+\sum_{s=0}^{-d_n}|\alpha_{s2}(K)|+\sum_{s=-1}^{-d_n}\sum_{t=2}^{2^s n}|\beta_{st}(K)|\right] \\
&\quad +\sum_{s=0}^{\infty}\frac{\lambda\ln n+s}{n}\sum_{t=1}^{2^s n}|\beta_{st}(K)| \\
&\leq \frac{\lambda\ln^{3/2}n}{n}\left[\sqrt{d_n+2}\sqrt{\frac{1}{2\pi}\int K^2(x)dx}+\ d_n\sqrt{\frac{1}{2\pi}\int K^2(x)dx}\right] \\
&\quad +\sum_{s=0}^{\infty}\frac{\lambda\ln n+s}{n}\ 2^{(-s+1)/2} \\
&= O\left(\frac{\ln^{5/2}n}{n}\right)\left(\sqrt{\int K^2(x)dx}+1\right) \qquad (20)
\end{aligned}$$

which completes (14). Two assertions are left to be verified.

Lemma I.4.1

$$\begin{aligned}
|\alpha_{01}(K)|+\sum_{s=0}^{d_n}|\alpha_{s2}(K)| &\leq \sqrt{d_n+2}\sqrt{\frac{1}{2\pi}\int K^2(x)dx} \\
\sum_{t=2}^{2^s n}|\beta_{st}(K)| &\leq \sqrt{\frac{1}{2\pi}\int K^2(x)dx} \qquad \text{for } s<0 \\
\sum_{t=1}^{2^s n}|\beta_{st}(K)| &\leq 2^{(-s+1)/2} \qquad \text{for } s\geq 0
\end{aligned}$$

Proof Since $\widehat{K}$, as well as all $\widehat{\varphi}_{st}$ are non-negative and the $\widehat{\varphi}_{st}$ are orthonormal, we can deduce:

$$\begin{aligned}
|\alpha_{01}(K)|+\sum_{s=0}^{d_n}|\alpha_{s2}(K)| &= \int\widehat{K}(\omega)\left[\widehat{\varphi}_{01}(\omega)+\sum_{s=0}^{d_n}\widehat{\varphi}_{s2}(\omega)\right]d\omega \\
&\leq \sqrt{\int\widehat{K}^2(\omega)d\omega}\sqrt{\int\widehat{\varphi}_{01}^2(\omega)d\omega+\sum_{s=0}^{d_n}\int\widehat{\varphi}_{s2}^2(\omega)d\omega} \\
&= \sqrt{\frac{1}{2\pi}\int\widehat{K}^2(\omega)d\omega}\ \sqrt{d_n+2}
\end{aligned}$$

May $\mathrm{TV}(\widehat{K}|\operatorname{supp}I_{st})$ denote the total variation of $\widehat{K}$ on the support of I_{st}, and the like for max and min. And let $\sum_t$ be $\sum_{t=2}^{2^s n}$ for $s<0$, and $\sum_{t=1}^{2^s n}$ for $s\geq 0$, respectively. Then for all $s\geq -d_n$ the following holds:

$$\sum_t|\beta_{st}(K)|$$

$$
\begin{aligned}
&= 2^{(s-1)/2}\sum_t \left|\int \widehat{K}(\omega)I_{s+1,2t-1}(\omega)d\omega - \int \widehat{K}(\omega)I_{s+1,2t}(\omega)d\omega\right| \\
&\le 2^{(s-1)/2}\sum_t \left|\max\left\{\widehat{K}\middle|\operatorname{supp} I_{s+1,2t-1}\right\}\int I_{s+1,2t-1}(\omega)d\omega - \min\left\{\widehat{K}\middle|\operatorname{supp} I_{s+1,2t}\right\}\int I_{s+1,2t}(\omega)d\omega\right| \\
&\le 2^{(s-1)/2}\sum_t 2^{-s}\ \mathrm{TV}\left(\widehat{K}\middle|\operatorname{supp} I_{st}\right) \\
&= 2^{-(s+1)/2}\ \mathrm{TV}\left(\widehat{K}\middle|\bigcup_t \operatorname{supp} Ist\right)
\end{aligned}
$$

For $s \ge 0$, the supports of the mother wavelets cover the whole interval $(-n, n)$, and we obtain $\mathrm{TV}(\widehat{K}|\bigcup \operatorname{supp} I_{st}) = \mathrm{TV}(\widehat{K}) \le 2$, due to unimodadility of $\widehat{K}$. For $s < 0$ we have the following short preliminary discussion: Because $\widehat{K}$ is monotonously decreasing on $[0, n)$ and symmetric around 0, it holds that $2|\omega|\widehat{K}^2(\omega) \le \int \widehat{K}^2$. Hence, for $s < 0$ and $\bigcup_t \operatorname{supp} I_{st} = (-n, -2^s] \cup [2^s, n)$:

$$
\begin{aligned}
\sum_t |\beta_{st}(K)| &\le 2^{-(s+1)/2}\ \mathrm{TV}\left(\widehat{K}\middle|\bigcup_t \operatorname{supp} I_{st}\right) \\
&\le 2^{-(s+1)/2}\ 2\,\widehat{K}\left(2^{-s}\right) \\
&\le 2^{-(s-1)/2}\ \frac{\sqrt{\int \widehat{K}^2(\omega)d\omega}}{\sqrt{2\cdot 2^{-s}}} \\
&\le \sqrt{\frac{1}{2\pi}\int K^2(x)dx} \qquad \square
\end{aligned}
$$

For the next lemma, it might be reminded that for densities f in a given Sobolev class $\mathcal{S}_\beta(L)$ with $\beta > 1/2$, the maximum of f is uniformly bounded.

Lemma I.4.2

$$
\begin{aligned}
P\left(\frac{1}{n(n-1)}\Big|\sum_{i\neq j} U_{\varphi_{st}}(X_i, X_j)\Big| > \frac{\lambda \ln^{3/2} n}{n}\right) &= O\left(n^{-\lambda^{2/3}}\right) \\
&\quad \text{for } (s,t) = (-d_n, 2), \ldots, (0,2), (0,1) \\
P\left(\frac{1}{n(n-1)}\Big|\sum_{i\neq j} U_{\psi_{st}}(X_i, X_j)\Big| > \frac{\lambda \ln^{3/2} n}{n}\right) &= O\left(n^{-\lambda^{2/3}}\right) \\
&\quad \text{for } s = -1, \ldots, -d_n; t = 2, \ldots, 2^s n \\
P\left(\frac{1}{n(n-1)}\Big|\sum_{i\neq j} U_{\psi_{st}}(X_i, X_j)\Big| > \frac{\lambda \ln n + s}{n}\right) &= O\left(n^{-\lambda^{2/3}} e^{-s}\right) \\
&\quad \text{for } s = 0, \ldots, \infty; t = 1, \ldots, 2^s n
\end{aligned}
$$

These bounds $O(.)$ are uniform in s and t, and as mentioned above, as well uniform over Sobolev classes with smoothness index greater than 1/2.

Proof From the Bernstein type inequality for degenerate U-statistics, shown by Arcones, Giné (1993) (see Appendix A.2), it follows that for all φ_{st}, and analogously for all ψ_{st} with $s<0$, there exist constants c_1 and c_2 independent from φ_{st} (and from ψ_{st} respectively), such that:

$$
\begin{aligned}
&P\left(\frac{1}{n(n-1)}\Big|\sum_{i\neq j} U_{\varphi_{st}}(X_i,X_j)\Big| > \frac{\lambda \ln^{3/2} n}{n}\right)\\
\leq\ & c_1 \exp\left\{-\frac{c_2(n-1)\frac{\lambda\ln^{3/2}n}{n}}{\sqrt{E|U_{\varphi_{st}}|^2}+\left(\frac{n-1}{n}\|\varphi_{st}\|_\infty^2\frac{\lambda\ln^{3/2}n}{n}\right)^{1/3}}\right\}\\
\leq\ & c_1 \exp\left\{-\frac{c_2(n-1)\frac{\lambda\ln^{3/2}n}{n}}{\|f\|_\infty^{1/2}\|\varphi_{st}\|_2+\left(\frac{n-1}{n}\frac{1}{(2\pi)^2}\|\widehat{\varphi}_{st}\|_1^2\frac{\lambda\ln^{3/2}n}{n}\right)^{1/3}}\right\}\\
=\ & c_1 \exp\left\{-\frac{c_2(n-1)\frac{\lambda\ln^{3/2}n}{n}}{\frac{1}{\sqrt{2\pi}}\|f\|_\infty^{1/2}\|\widehat{\varphi}_{st}\|_2+\left(\frac{n-1}{(2\pi)^2 n}\,2^{-(s-1)}\,\frac{\lambda\ln^{3/2}n}{n}\right)^{1/3}}\right\}\\
\leq\ & c_1 \exp\left\{-\frac{c_2(n-1)\frac{\lambda\ln^{3/2}n}{n}}{\frac{1}{\sqrt{2\pi}}\|f\|_\infty^{1/2}+\left(\frac{n-1}{(2\pi)^2 n}\,2^{d_n+1}\,\frac{\lambda\ln^{3/2}n}{n}\right)^{1/3}}\right\}\\
=\ & c_1 \exp\left\{-\frac{c_2\frac{n-1}{n}\lambda\ln^{3/2}n}{\frac{1}{\sqrt{2\pi}}\|f\|_\infty^{1/2}+\left(\frac{n-1}{(2\pi)^2 n}\,2n\,\frac{\lambda\ln^{3/2}n}{n}\right)^{1/3}}\right\}\\
=\ & c_1 \exp\left\{-\frac{c_2\left(\frac{n-1}{n}\right)^{2/3}\lambda^{2/3}\ln n}{\frac{1}{\sqrt{2\pi}}\|f\|_\infty^{1/2}\left(\frac{\lambda(n-1)}{n}\right)^{-1/3}\ln^{-1/2}n+2^{-1/3}\pi^{-2/3}}\right\}\\
=\ & O\left(\exp\left\{-\frac{\lambda^{2/3}\ln n}{1+\|f\|_\infty^{1/2}\ln^{-1/2}n}\right\}\right)
\end{aligned}
$$

which is a uniform $O(n^{-\lambda^{2/3}})$ for densities in $\mathcal{S}_\beta(L)$, $\beta>1/2$.
For ψ_{st} with $s\geq 0$, approximate $2^{-s}n^{-1}(\lambda\ln n+s)$ by $\lambda+1/2$, such that:

$$
\begin{aligned}
&P\left(\frac{1}{n(n-1)}\Big|\sum_{i\neq j} U_{\psi_{st}}(X_i,X_j)\Big| > \frac{\lambda \ln n+s}{n}\right)\\
\leq\ & c_1 \exp\left\{-\frac{c_2(n-1)\frac{\lambda\ln n+s}{n}}{\sqrt{E|U_{\psi_{st}}|^2}+\left(\frac{n-1}{n}\|\psi_{st}\|_\infty^2\frac{\lambda\ln n+s}{n}\right)^{1/3}}\right\}\\
\leq\ & c_1 \exp\left\{-\frac{c_2(n-1)\frac{\lambda\ln n+s}{n}}{\frac{1}{\sqrt{2\pi}}\|f\|_\infty^{1/2}+\left(\frac{n-1}{n}\frac{1}{(2\pi)^2}\,2^{-(s-1)}\frac{\lambda\ln n+s}{n}\right)^{1/3}}\right\}\\
\leq\ & c_1 \exp\left\{-\frac{c_2(n-1)\frac{\lambda\ln n+s}{n}}{\frac{1}{\sqrt{2\pi}}\|f\|_\infty^{1/2}+\left(\frac{n-1}{n}\frac{2\lambda+1}{(2\pi)^2}\right)^{1/3}}\right\}
\end{aligned}
$$

$$= O\left(\exp\left\{-\frac{\lambda^{2/3}\ln n + \lambda^{-1/3}s}{\|f\|_\infty^{1/2}}\right\}\right)$$

a uniform $O(n^{-\lambda^{2/3}}e^{-s})$ for densities in $\mathcal{S}_\beta(L)$, $\beta > 1/2$. □

A_{n1}^c is then the union of all complementary sets and its probability decreases like $n^{-\lambda^{2/3}+1}$:

$$\begin{aligned}
P\left(A_{n1}^c\right) &\leq P\left(\frac{1}{n(n-1)}\Big|\sum_{i\neq j} U_{\varphi_{01}}(X_i,X_j)\Big| > \frac{\lambda\ln^{3/2}n}{n}\right) \\
&+ \sum_{s=0}^{-d_n} P\left(\frac{1}{n(n-1)}\Big|\sum_{i\neq j} U_{\varphi_{s2}}(X_i,X_j)\Big| > \frac{\lambda\ln^{3/2}n}{n}\right) \\
&+ \sum_{s=-1}^{-d_n}\sum_{t=2}^{2^s n} P\left(\frac{1}{n(n-1)}\Big|\sum_{i\neq j} U_{\psi_{st}}(X_i,X_j)\Big| > \frac{\lambda\ln^{3/2}n}{n}\right) \\
&+ \sum_{s=0}^{\infty}\sum_{t=1}^{2^s n} P\left(\frac{1}{n(n-1)}\Big|\sum_{i\neq j} U_{\psi_{st}}(X_i,X_j)\Big| > \frac{\lambda\ln n + s}{n}\right) \\
&= O\left(n^{-\lambda^{2/3}}\right)\left[1 + \sum_{s=0}^{-d_n} 1 + \sum_{s=-1}^{-d_n}\sum_{t=2}^{2^s n} 1\right] + \sum_{s=0}^{\infty}\sum_{t=1}^{2^s n} O\left(n^{-\lambda^{2/3}}e^{-s}\right) \\
&= O\left(n^{-\lambda^{2/3}}\right) O(\ln n + n) + O\left(n^{-\lambda^{2/3}}\right) O(n) \qquad (21)
\end{aligned}$$

4.2 The empirical process

We are now going to consider the bias term in the difference $\widetilde{CV}(K) - ISE(K)$ for kernel functions K in $\mathcal{K}_n$:

$$\frac{1}{n}\sum_{j=1}^{n}\Big(b_K(X_j) + h_f(X_j)\Big) - E\Big[b_K(X_j) + h_f(X_j)\Big]$$

where $b_K(x) = f * K(x) - f(x)$ and h_f is the high-frequency component of f (definition (11)). Because of the bounded support of $\widehat{K}$, $b_k + h_f$ takes the form:

$$\begin{aligned}
b_K(x) + h_f(x) &= f * K(x) - f(x) + h_f(x) \\
&= \frac{1}{2\pi}\int \widehat{f}(\omega)\Big(\widehat{K}(\omega) - 1\Big)e^{-i\omega x}d\omega + \frac{1}{2\pi}\int \widehat{f}(\omega)\Big(1 - I_n(\omega)\Big)e^{-i\omega x}d\omega \\
&= \frac{1}{2\pi}\int \left[\widehat{f}(\omega)\Big(\widehat{K}(\omega)I_n(\omega) - 1\Big) + \widehat{f}(\omega)\Big(1 - I_n(\omega)\Big)\right] e^{-i\omega x}d\omega \\
&= \frac{1}{2\pi}\int \widehat{f}(\omega)\Big(\widehat{K}(\omega) - 1\Big)I_n(\omega)e^{-i\omega x}d\omega \qquad (22)
\end{aligned}$$

the low-frequency component of the bias and exactly that part which really depends on the kernel. Again we want to apply an additive decomposition in the Fourier domain, this time to the function $\widehat{b}_K + \widehat{h}_f = \widehat{b}_K \cdot I_n$.

For a new function basis, this time dependent of $\widehat{f}$, let us define the integral of $|\widehat{f}|^2$ over $(-\omega, \omega)$ as a function $F(\omega)$.

$$F(\omega) := \int_{-\omega}^{\omega} |\widehat{f}(\tau)|^2 d\tau$$

This map transforms the ω-halfaxis $[0,\infty)$ by mapping $\omega \longmapsto F(\omega)$ to the interval $[0, \|\widehat{f}\|_2^2)$.

$$F(0)=0, \quad F_n := F(n) = \int_{-n}^{n} |\widehat{f}(\tau)|^2 d\tau, \quad \lim_{\omega\to\infty} F(\omega) = \|\widehat{f}\|_2^2$$

The initial value of an interval, say $[2^{-s}(t-1)F_n, 2^{-s}tF_n)$ with length 2^{-s}, on this axis is the interval $[F^{-1}(2^{-s}(t-1)F_n), F^{-1}(2^{-s}tF_n))$ on the original axis. The integral of $|\widehat{f}|^2$ over the initial interval is obviously $\frac{1}{2}2^{-s}F_n$.

$$2^{-s}F_n = \int_{-F^{-1}(2^{-s}tF_n)}^{F^{-1}(2^{-s}tF_n)} |\widehat{f}(\omega)|^2 d\omega - \int_{-F^{-1}(2^{-s}(t-1)F_n)}^{F^{-1}(2^{-s}(t-1)F_n)} |\widehat{f}(\omega)|^2 d\omega$$

Define the indicator functions:

$$I'_{st}(\omega) \;:=\; I\left(|\omega| \in \left[F^{-1}\left(2^{-s}(t-1)F_n\right), F^{-1}\left(2^{-s}tF_n\right)\right)\right) \tag{23}$$

satisfying $\int |\widehat{f}(\omega)|^2 I'_{st}(\omega)d\omega = 2^{-s}F_n$, and the orthonomal wavelet functions:

$$\begin{aligned}
\widehat{\varphi}'_{st}(\omega) \;&:=\; 2^{s/2}F_n^{-1/2}\widehat{f}(\omega)I'_{st}(\omega), \qquad \text{for} \quad s=1,\dots,s_n-1 \text{ with } t=2^s-1 \text{ and} \\
&\qquad\qquad s=s_n, t=2^{s_n}-1, 2^{s_n} \text{ where } s_n := \lceil \ln n/\ln 2 \rceil \\
\widehat{\psi}'_{st}(\omega) \;&:=\; 2^{s/2}F_n^{-1/2}\widehat{f}(\omega)\left[I'_{s+1,2t-1}(\omega) - I'_{s+1,2t}(\omega)\right], \quad \text{for } s=1,\dots,s_n-1 \text{ with} \\
&\qquad\qquad t=1,\dots,2^s-1 \text{ and } s=s_n,\dots,\infty \text{ with} \\
&\qquad\qquad t=1,\dots,2^s
\end{aligned} \tag{24}$$

$\{\widehat{\varphi}'_{st}|s=1,\dots s_n-1, t=2^s-1\} \cup \{\widehat{\varphi}'_{s_n,2^{s_n}}\} \cup \{\widehat{\psi}'_{st}|s=1,\dots,s_n-1, t=1,\dots 2^s-1\} \cup \{\widehat{\psi}'_{st}|s \geq s_n, t=1,\dots 2^s\}$ represent a complete orthonormal basis for the set of all functions $\{\widehat{f}\cdot\widehat{g}\cdot I_n \,|\, \widehat{g} \in L_2\}$, which the bias functions $\widehat{b}_K \cdot I_n$ belong to for all $K \in \mathcal{K}$. With decomposition on this basis, we can write $\widehat{b}_K \cdot I_n$ as

$$\begin{aligned}
\widehat{b}_K(\omega)I_n(\omega) \;&=\; \sum_{s=1}^{s_n} \alpha'_{s2^s-1}(b_K)\widehat{\varphi}'_{s2^s-1}(\omega) + \alpha'_{s_n2^{s_n}}(b_K)\widehat{\varphi}'_{s_n2^{s_n}}(\omega) + \sum_{s=1}^{s_n-1}\sum_{t=1}^{2^s-1} \beta'_{st}(b_K)\widehat{\psi}'_{st}(\omega) \\
&\quad + \sum_{s=s_n}^{\infty}\sum_{t=1}^{2^s} \beta'_{st}(b_K)\widehat{\psi}'_{st}(\omega)
\end{aligned} \tag{25}$$

$$\alpha'_{st}(b_K) \;:=\; \int \widehat{b}_K(\omega)\overline{\widehat{\varphi}'}_{st}(\omega)d\omega \qquad \text{and} \qquad \beta'_{st}(b_K) \;:=\; \int \widehat{b}_K(\omega)\overline{\widehat{\psi}'}_{st}(\omega)d\omega$$

The inverse Fourier transform yields:

$$\begin{aligned}
b_K(x) + h_f(x) \;&=\; \sum_{s=1}^{s_n} \alpha'_{s2^s-1}(b_K)\varphi'_{s2^s-1}(x) + \alpha'_{s_n2^{s_n}}(b_K)\varphi'_{s_n2^{s_n}}(x) + \sum_{s=1}^{s_n-1}\sum_{t=1}^{2^s-1} \beta'_{st}(b_K)\psi'_{st}(x) \\
&\quad + \sum_{s=s_n}^{\infty}\sum_{t=1}^{2^s} \beta'_{st}(b_K)\psi'_{st}(x)
\end{aligned}$$

which gives in turn

$$\frac{1}{n}\sum_{j=1}^{n}\Big(b_K(X_j) + h_f(X_j)\Big) - E\Big[b_K(X_j) + h_f(X_j)\Big]$$

$$= \sum_{s=1}^{s_n} \alpha'_{s2^s-1}(b_K)\left[\frac{1}{n}\sum_{j=1}^{n}\varphi'_{s2^s-1}(X_j) - E\varphi'_{s2^s-1}(X_j)\right] + \alpha'_{s_n 2^{s_n}}(b_K)\left[\frac{1}{n}\sum_{j=1}^{n}\varphi'_{s_n 2^{s_n}}(X_j) - E\varphi'_{s_n 2^{s_n}}(X_j)\right]$$
$$+ \sum_{s=1}^{s_n-1}\sum_{t=1}^{2^s-1}\beta'_{st}(b_K)\left[\frac{1}{n}\sum_{j=1}^{n}\psi'_{st}(X_j) - E\psi'_{st}(X_j)\right] + \sum_{s=s_n}^{\infty}\sum_{t=1}^{2^s}\beta'_{st}(b_K)\left[\frac{1}{n}\sum_{j=1}^{n}\psi'_{st}(X_j) - E\psi'_{st}(X_j)\right] \quad (26)$$

Again, we will proceed separately with the aim of finding bounds to the deterministic wavelet coefficients and the stochastic processes. Lemma I.4.3 shows that

$$\begin{aligned}
\sum_{s=1}^{s_n}|\alpha'_{s2^s-1}(b_K)| + |\alpha'_{s_n 2^{s_n}}(b_K)| &\leq \sqrt{s_n+1}\,\sqrt{\frac{1}{2\pi}\int b_K^2(x)dx} \\
\sum_{t=1}^{2^s-1}|\beta'_{st}(b_K)| &\leq \sqrt{\frac{2}{\pi}\int b_K^2(x)dx} \qquad \text{for } s < s_n \\
\sum_{t=1}^{2^s}|\beta'_{st}(b_K)| &\leq 2\cdot 2^{-s/2}\|f\|_2 \qquad \text{for } s \geq s_n
\end{aligned}$$

Over a set of "favorable events", whose complement has an asymptotically decreasing probability (Lemma I.4.4 and inequality (29)), the empirical processes can be controlled. For $\lambda < \infty$

$$\begin{aligned}
A_{n2} := \Big\{(X_1,\dots,X_n):\ & \tfrac{1}{n}|\textstyle\sum_{j=1}^{n}\varphi'_{st}(X_j) - E\varphi'_{st}(X_j)| \leq \tfrac{\lambda\ln n}{\sqrt{n}}, \quad (s,t) = (1,1),\dots,(s_n, 2^{s_n}-1),(s_n,2^{s_n}); \\
& \tfrac{1}{n}|\textstyle\sum_{j=1}^{n}\psi'_{st}(X_j) - E\psi'_{st}(X_j)| \leq \tfrac{\lambda\ln n}{\sqrt{n}}, \quad s = 1,\dots,s_n-1, t = 1,\dots,2^s-1; \\
& \tfrac{1}{n}|\textstyle\sum_{j=1}^{n}\psi'_{st}(X_j) - E\psi'_{st}(X_j)| \leq \tfrac{\lambda\ln n + s}{\sqrt{n}}, \quad s \geq s_n, t = 1,\dots,2^s\Big\}
\end{aligned} \quad (27)$$

$$\text{and} \qquad P(A_{n2}^c) = O\Big(n^{-\lambda+1}\Big)$$

uniformly over Sobolev classes with smoothness index $> 1/2$. Following (26), taking into account that $2^{s_n} \leq n^{-1}$, it holds on A_{n2}:

$$\begin{aligned}
&\frac{1}{n}\Big|\sum_{j=1}^{n}\Big(b_K(X_j) + h_f(X_j)\Big) - E\Big[b_K(X_j) + h_f(X_j)\Big]\Big| \\
\leq\ & \frac{\lambda\ln n}{\sqrt{n}}\left[\sum_{s=1}^{s_n}\left|\alpha'_{s2^s-1}(b_K)\right| + \left|\alpha'_{s_n2^{s_n}}(b_K)\right| + \sum_{s=1}^{s_n-1}\sum_{t=1}^{2^s-1}\left|\beta'_{st}(b_K)\right|\right] + \sum_{s=s_n}^{\infty}\frac{\lambda\ln n + s}{\sqrt{n}}\sum_{t=1}^{2^s}\left|\beta'_{st}(b_K)\right| \\
\leq\ & \frac{\lambda\ln n}{\sqrt{n}}\left[\sqrt{s_n+1}\,\sqrt{\frac{1}{2\pi}\int b_K^2(x)dx} + (s_n-1)\sqrt{\frac{2}{\pi}\int b_K^2(x)dx}\right] + 2\,\|f\|_2\sum_{s=s_n}^{\infty}\frac{\lambda\ln n + s}{\sqrt{n}}2^{-s/2} \\
=\ & O\left(\frac{\ln^2 n}{\sqrt{n}}\right)\sqrt{\int b_K^2(x)dx} + \|f\|_2\,\frac{\ln n}{\sqrt{n}}\,O\left(2^{-s_n/2}\right) \\
=\ & O\left(\frac{\ln^2 n}{\sqrt{n}}\right)\sqrt{\int b_K^2(x)dx} + \|f\|_2\,\frac{\ln n}{\sqrt{n}}\,O\left(n^{-1/2}\right) \\
=\ & O\left(\frac{\ln^2 n}{\sqrt{n}}\right)\left(\sqrt{\int b_K^2(x)dx} + \frac{\|f\|_2}{\sqrt{n}}\right)
\end{aligned} \quad (28)$$

which completes (15). Now we proof the assertion on the wavelet coefficients.

Lemma I.4.3

$$
\begin{aligned}
\sum_{s=1}^{s_n} |\alpha'_{s2^s-1}(b_K)| + |\alpha'_{s_n 2^{s_n}}(b_K)| &\leq \sqrt{s_n+1}\sqrt{\frac{1}{2\pi}\int b_K^2(x)dx} \\
\sum_{t=1}^{2^s-1} |\beta'_{st}(b_K)| &\leq \sqrt{\frac{2}{\pi}\int b_K^2(x)dx} \qquad \text{for } s < s_n \\
\sum_{t=1}^{2^s} |\beta'_{st}(b_K)| &\leq 2\cdot 2^{-s/2}\|f\|_2 \qquad \text{for } s \geq s_n
\end{aligned}
$$

Proof The father wavelet coefficients are bounded in the same way as in Lemma I.4.1, such that

$$
\begin{aligned}
\sum_{s=1}^{s_n} |\alpha'_{s2^s-1}(b_K)| + |\alpha'_{s_n 2^{s_n}}(b_K)| &= \sum_{s=1}^{s_n}\Big|\int \widehat{b}_K(\omega)\overline{\widehat{\varphi}'}_{s2^s-1}(\omega)d\omega\Big| + \Big|\int \widehat{b}_K(\omega)\overline{\widehat{\varphi}'}_{s_n 2^{s_n}}(\omega)d\omega\Big| \\
&\leq \int |\widehat{b}_K(\omega)| \left[\sum_{s=1}^{s_n} |\widehat{\varphi}'_{s2^s-1}(\omega)| + |\widehat{\varphi}'_{s_n 2^{s_n}}(\omega)|\right] d\omega \\
&\leq \sqrt{\int |\widehat{b}_K(\omega)|^2 d\omega}\sqrt{\sum_{s=1}^{s_n}\int |\widehat{\varphi}'_{s2^s-1}(\omega)|^2 d\omega + \int |\widehat{\varphi}'_{s_n 2^{s_n}}(\omega)|^2 d\omega} \\
&= \sqrt{\frac{1}{2\pi}\int |b_K(x)|^2 dx}\,\sqrt{s_n+1}
\end{aligned}
$$

For $\sum_t$ take the summation $\sum_{t=1}^{2^s-1}$ when $s < s_n$ and $\sum_{t=1}^{2^s}$ when $s \geq s_n$. For every t in the summation range, choose an arbitrary $\omega_{st} \in [F^{-1}(2^{-s}(t-1)F_n), F^{-1}(2^{-s}tF_n))$. Again let $\mathrm{TV}(\widehat{K}|\mathrm{supp}\, I'_{st})$ be the total variation of $\widehat{K}$ over the support of I'_{st}.

$$
\begin{aligned}
\sum_t |\beta'_{st}(b_K)| &= \sum_t \Big|\int \widehat{b}_K(\omega)\overline{\widehat{\psi}'}_{st} d\omega\Big| \\
&= \sum_t 2^{s/2}F_n^{-1/2}\Big|\int \widehat{f}(\omega)\left(1-\widehat{K}(\omega)\right)\overline{\widehat{f}}(\omega)\left[I'_{s+1,2t-1}(\omega) - I'_{s+1,2t}(\omega)\right]d\omega\Big| \\
&= \sum_t 2^{s/2}F_n^{-1/2}\Big|\int |\widehat{f}(\omega)|^2\left(1-\widehat{K}(\omega_{st})\right)\left[I'_{s+1,2t-1}(\omega) - I'_{s+1,2t}(\omega)\right]d\omega \\
&\quad + \int |\widehat{f}(\omega)|^2\left(\widehat{K}(\omega_{st}) - \widehat{K}(\omega)\right)\left[I'_{s+1,2t-1}(\omega) - I'_{s+1,2t}(\omega)\right]d\omega\Big| \\
&= \sum_t 2^{s/2}F_n^{-1/2}\Big|\left(1-\widehat{K}(\omega_{st})\right)\int |\widehat{f}(\omega)|^2\left[I'_{s+1,2t-1}(\omega) - I'_{s+1,2t}(\omega)\right]d\omega \\
&\quad + \int |\widehat{f}(\omega)|^2\left(\widehat{K}(\omega_{st}) - \widehat{K}(\omega)\right) I'_{s+1,2t-1}(\omega)d\omega \\
&\quad - \int |\widehat{f}(\omega)|^2\left(\widehat{K}(\omega_{st}) - \widehat{K}(\omega)\right) I'_{s+1,2t}(\omega)d\omega\Big|
\end{aligned}
$$

$$
\begin{aligned}
&\le \sum_t 2^{s/2} F_n^{-1/2}\left[0+\int |\widehat{f}(\omega)|^2 \ \mathrm{TV}\left(\widehat{K}\Big|\operatorname{supp} I'_{s+1,2t-1}\right) I'_{s+1,2t-1}(\omega)d\omega\right.\\
&\qquad \left. + \int |\widehat{f}(\omega)|^2 \ \mathrm{TV}\left(\widehat{K}\Big|\operatorname{supp} I'_{s+1,2t}(|\omega|)\right) I'_{s+1,2t}(\omega)d\omega\right]\\
&= \sum_t 2^{s/2} F_n^{-1/2} \ \mathrm{TV}\left(\widehat{K}\Big|\operatorname{supp} I'_{st}\right)\int |\widehat{f}(\omega)|^2 I'_{st}(\omega)d\omega\\
&= \sum_t 2^{s/2} F_n^{-1/2} \ \mathrm{TV}\left(\widehat{K}\Big|\operatorname{supp} I'_{st}\right) F_n 2^{-s}\\
&= 2^{-s/2} F_n^{1/2} \ \mathrm{TV}\left(\widehat{K}\Big|\operatorname{supp} \bigcup_t I'_{st}\right)
\end{aligned}
$$

The mother wavelets on the scales $s \ge s_n$ are defined over the whole $(-n, n)$, therefore $\mathrm{TV}(\widehat{K}|\operatorname{supp}\bigcup I'_{st}) = \mathrm{TV}(\widehat{K}) \le 2$. For $s < s_n$, the support of the mother wavelets is $[-F^{-1}((1-2^s)F_n), F^{-1}((1-2^s)F_n))$. On this support, the total variation amounts to at most $2[1-\widehat{K}\left(F^{-1}((1-2^s)F_n)\right)]$.

$$
\begin{aligned}
\sum_t |\beta'_{st}(b_K)| &\le 2^{-s/2} F_n^{1/2} 2\left[1-\widehat{K}\left(F^{-1}\Big((1-2^s)F_n\Big)\right)\right]\\
&= 2^{-s/2} F_n^{1/2}\left(\int |\widehat{\varphi}'_{s2^s}(\omega)|^2 d\omega\right) 2\left[1-\widehat{K}\left(F^{-1}\Big((1-2^s)F_n\Big)\right)\right]\\
&= 2^{s/2} F_n^{-1/2}\int |\widehat{f}(\omega)|^2 I'_{s2^s}(\omega)d\omega \ 2\left[1-\widehat{K}\left(F^{-1}\Big((1-2^s)F_n\Big)\right)\right]\\
&\le 2^{s/2} F_n^{-1/2} \ 2\int \widehat{f}(\omega)\left[1-\widehat{K}(\omega)\right]\overline{\widehat{f}}(\omega) I'_{s2^s}(\omega)d\omega\\
&= 2\int \widehat{b}_K(\omega) \ \overline{\widehat{\varphi}'}_{s2^s}(\omega)d\omega\\
&\le 2\sqrt{\int |\widehat{b}_K(\omega)|^2 d\omega}\sqrt{\int |\widehat{\varphi}'_{s2^s}(\omega)|^2 d\omega}\\
&= 2\sqrt{\frac{1}{2\pi}\int b_K^2(x)dx} \qquad \square
\end{aligned}
$$

Lemma I.4.4 The following inequalities hold uniformly for all indicated s and t, and furthermore uniformly for $f \in \mathcal{S}_\beta(L)$, $\beta > 1/2$:

$$
\begin{aligned}
P\left(\frac{1}{n}\Big|\sum_{j=1}^n \varphi'_{st}(X_j) - E\varphi'_{st}(X_j)\Big| > \frac{\lambda \ln n}{\sqrt{n}}\right) &= O\Big(n^{-\lambda}\Big), \quad (s,t) = (1,1),\ldots,(s_n, 2^{s_n}-1),\\
&\qquad\qquad \text{and } (s_n, 2^{s_n})\\
P\left(\frac{1}{n}\Big|\sum_{j=1}^n \psi'_{st}(X_j) - E\psi'_{st}(X_j)\Big| > \frac{\lambda \ln n}{\sqrt{n}}\right) &= O\Big(n^{-\lambda}\Big), \quad s = 1,\ldots,s_{n-1} \text{ and}\\
&\qquad\qquad t = 1,\ldots,2^s-1\\
P\left(\frac{1}{n}\Big|\sum_{j=1}^n \psi'_{st}(X_j) - E\psi'_{st}(X_j)\Big| > \frac{\lambda \ln n + s}{\sqrt{n}}\right) &= O\Big(n^{-\lambda}e^{-s}\Big), \quad s = s_n,\ldots,\infty \text{ and}\\
&\qquad\qquad t = 1,\ldots,2^s
\end{aligned}
$$

Proof According to Bernstein's inequality (e.g. Shorack, Wellner (1986), p. 855), for all φ'_{st} and analogously for all ψ'_{st} with $s < s_n$ it holds that:

$$\begin{aligned}
& P\left(\frac{1}{n}\Big|\sum_{j=1}^{n}\varphi'_{st}(X_j) - E\varphi'_{st}(X_j)\Big| > \frac{\lambda\ln n}{\sqrt{n}}\right) \\
\leq \;& 2\exp\left\{-\frac{\frac{n}{2}\left(\frac{\lambda\ln n}{\sqrt{n}}\right)^2}{\sqrt{E|\varphi'_{st}|^2} + \|\varphi'_{st}\|_\infty\frac{\lambda\ln n}{3\sqrt{n}}}\right\} \\
\leq \;& 2\exp\left\{-\frac{\lambda^2\ln^2 n}{2\|f\|_\infty\|\varphi'_{st}\|_2^2 + \frac{1}{2\pi}\|\widehat{\varphi}'_{st}\|_1\frac{2\lambda\ln n}{3\sqrt{n}}}\right\} \\
\leq \;& 2\exp\left\{-\frac{\lambda^2\ln^2 n}{\frac{1}{\pi}\|f\|_\infty\|\widehat{\varphi}'_{st}\|_2^2 + \frac{1}{\pi}\|\widehat{\varphi}'_{st}\|_2\|I'_{st}\|_2\frac{\lambda\ln n}{3\sqrt{n}}}\right\} \\
\leq \;& 2\exp\left\{-\frac{\lambda^2\ln^2 n}{\frac{1}{\pi}\|f\|_\infty + \frac{1}{\pi}\sqrt{2n}\,\frac{\lambda\ln n}{3\sqrt{n}}}\right\} \\
\leq \;& O\left(\exp\left\{-\frac{\lambda\ln n}{1+\|f\|_\infty\lambda^{-1}\ln^{-1}n}\right\}\right)
\end{aligned}$$

which is a uniform $O(n^{-\lambda})$ for $f \in \mathcal{S}_\beta(L)$, $\beta > 1/2$. For ψ'_{st} with $s \geq s_n = \lceil \ln n/\ln 2\rceil$:

$$\begin{aligned}
& P\left(\frac{1}{n}\Big|\sum_{j=1}^{n}\psi'_{st}(X_j) - E\psi'_{st}(X_j)\Big| > \frac{\lambda\ln n + s}{\sqrt{n}}\right) \\
\leq \;& 2\exp\left\{-\frac{\frac{n}{2}\left(\frac{\lambda\ln n+s}{\sqrt{n}}\right)^2}{\sqrt{E|\psi'_{st}|^2} + \|\psi'_{st}\|_\infty\frac{\lambda\ln n+s}{3\sqrt{n}}}\right\} \\
\leq \;& 2\exp\left\{-\frac{(\lambda\ln n + s)^2}{\frac{1}{\pi}\|f\|_\infty + \frac{1}{\pi}\sqrt{2n}\,\frac{\lambda\ln n+s}{3\sqrt{n}}}\right\} \\
= \;& 2\exp\left\{-\frac{\lambda\ln n + s}{\frac{1}{\pi}\|f\|_\infty\,(\lambda\ln n + s)^{-1} + \frac{\sqrt{2}}{3\pi}}\right\} \\
\leq \;& 2\exp\left\{-\frac{\lambda\ln n + s}{\frac{1}{\pi}\|f\|_\infty\,(\lambda\ln n + \lceil\ln n/\ln 2\rceil)^{-1} + \frac{\sqrt{2}}{3\pi}}\right\} \\
= \;& O\left(\exp\left\{-\frac{\lambda\ln n + s}{1+\|f\|_\infty\lambda^{-1}\ln^{-1}n}\right\}\right)
\end{aligned}$$

a uniform $O(n^{-\lambda}e^{-s})$ for $f \in \mathcal{S}_\beta(L)$, $\beta > 1/2$. □

$$\begin{aligned}
P\left(A_{n2}^c\right) \;\leq\; & \sum_{s=1}^{s_n} P\left(\frac{1}{n}\Big|\sum_{j=1}^{n}\varphi'_{s2^s-1}(X_j) - E\varphi'_{s2^s-1}(X_j)\Big| > \frac{\lambda\ln n}{\sqrt{n}}\right) \\
& + P\left(\frac{1}{n}\Big|\sum_{j=1}^{n}\varphi'_{s_n 2^{s_n}}(X_j) - E\varphi'_{s_n 2^{s_n}}(X_j)\Big| > \frac{\lambda\ln n}{\sqrt{n}}\right)
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{s=1}^{s_n-1}\sum_{t=1}^{2^s-1} P\left(\frac{1}{n}\Big|\sum_{j=1}^{n}\psi'_{st}(X_j)-E\psi'_{st}(X_j)\Big| > \frac{\lambda\ln n}{\sqrt{n}}\right)\\
&+\sum_{s=s_n}^{\infty}\sum_{t=1}^{2^s} P\left(\frac{1}{n}\Big|\sum_{j=1}^{n}\psi'_{st}(X_j)-E\psi'_{st}(X_j)\Big| > \frac{\lambda\ln n+s}{\sqrt{n}}\right)\\
=\;& O\left(n^{-\lambda}\right)\left[\sum_{s=1}^{s_n}1+1+\sum_{s=1}^{s_n-1}\sum_{t=1}^{2^s-1}1\right]+\sum_{s=s_n}^{\infty}\sum_{t=1}^{2^s}O\left(n^{-\lambda}e^{-s}\right)\\
=\;& O\left(n^{-\lambda}\right)O(\ln n+n)+O\left(n^{-\lambda}\right) \qquad (29)
\end{aligned}
$$

so $P(A_{n2}^c)$ is less than an $O(n^{-\lambda+1})$.

5 ISE-MISE

It is possible to bound ISE-MISE in the same way as CV-ISE in Section 4, even on the same set of "favorable events" A_n. For this purpose, let us split ISE-MISE into an empirical degenerate U-process and an empirical process:

$$
\begin{aligned}
ISE(K)-MISE(K) &= \int\left(\widetilde{f}_K(x)-f(x)\right)^2dx - E\int\left(\widetilde{f}_K(x)-f(x)\right)^2dx\\
&= \int\left(\widetilde{f}_K(x)-E\widetilde{f}_K(x)+E\widetilde{f}_K(x)-f(x)\right)^2dx\\
&\quad-\left[\int\left(E\widetilde{f}_K(x)-f(x)\right)^2dx+E\int\left(\widetilde{f}_K(x)-E\widetilde{f}_K(x)\right)^2dx\right]\\
&= \left[\int\left(\widetilde{f}_K(x)-E\widetilde{f}_K(x)\right)^2dx-E\int\left(\widetilde{f}_K(x)-E\widetilde{f}_K(x)\right)^2dx\right]\\
&\quad+2\int\left(E\widetilde{f}_K(x)-f(x)\right)\left(\widetilde{f}_K(x)-E\widetilde{f}_K(x)\right)dx \qquad (30)
\end{aligned}
$$

The latter term on the right-hand side can be reordered, such that with $b_K = E\widetilde{f}_K - f$:

$$
\begin{aligned}
\int\Big(E\widetilde{f}_K(x)-f(x)\Big)\Big(\widetilde{f}_K(x)-E\widetilde{f}_K(x)\Big)dx &= \frac{1}{n}\sum_{j=1}^{n}\int b_K(x)K(X_j-x)dx - E\int b_K(x)K(X_j-x)dx\\
&= \frac{1}{n}\sum_{j=1}^{n}\left(b_K*K\right)(X_j)-E\left[\left(b_K*K\right)(X_j)\right] \qquad (31)
\end{aligned}
$$

$\widehat{b_k*K}=\widehat{b}_K\cdot\widehat{K}$ is 0 outside of $(-n,n)$, because of the multiplication by $\widehat{K}$. Therefore it can be decomposed using the wavelet basis for b_K+h_j constructed in Subsection 4.2.

$$
\alpha'_{st}(b_K*K) \;:=\; \int\widehat{b}_K(\omega)\widehat{K}(\omega)\overline{\widehat{\varphi}'}_{st}(\omega)d\omega \quad\text{and}\quad \beta'_{st}(b_K*K) \;:=\; \int\widehat{b}_K(\omega)\widehat{K}(\omega)\overline{\widehat{\psi}'}_{st}(\omega)d\omega
$$

The set of "favorable events" is A_{n2} as for the empirical process of b_K+h_f, on which it can be shown in the same way as in Lemma I.4.3 that:

$$
\frac{1}{n}\Big|\sum_{j=1}^{n}b_K*K(X_j)-E\left[(b_K*K)(X_j)\right]\Big|
$$

$$
\begin{aligned}
&= \Big|\sum_{s=1}^{s_n} \alpha'_{s2^s-1}(b_K * K)\left[\frac{1}{n}\sum_{j=1}^{n}\varphi'_{s2^s-1}(X_j) - E\varphi'_{s2^s-1}(X_j)\right] \\
&\quad + \alpha'_{s_n2^{s_n}}(b_K * K)\left[\frac{1}{n}\sum_{j=1}^{n}\varphi'_{s_n2^{s_n}}(X_j) - E\varphi'_{s_n2^{s_n}}(X_j)\right] \\
&\quad + \sum_{s=1}^{s_n-1}\sum_{t=1}^{2^s-1}\beta'_{st}(b_K * K)\left[\frac{1}{n}\sum_{j=1}^{n}\psi'_{st}(X_j) - E\psi'_{st}(X_j)\right] \\
&\quad + \sum_{s=s_n}^{\infty}\sum_{t=1}^{2^s}\beta'_{st}(b_K * K)\left[\frac{1}{n}\sum_{j=1}^{n}\psi'_{st}(X_j) - E\psi'_{st}(X_j)\right]\Big| \\
&= O\left(\frac{\ln^2 n}{\sqrt{n}}\right)\left(\sqrt{\int |\widehat{b}_K(\omega)\widehat{K}(\omega)|^2 d\omega} \;+\; \frac{\|f\|_2 \mathrm{TV}(\widehat{K}(1-\widehat{K}))}{\sqrt{n}}\right) \\
&= O\left(\frac{\ln^2 n}{\sqrt{n}}\right)\left(\sqrt{\int b_K^2(x)dx} \;+\; \frac{\|f\|_2}{\sqrt{n}}\right) \qquad (32)
\end{aligned}
$$

Since $\widehat{K}$ is non-negative and monotonously decreasing on $\mathbb{R}^+$, so is $\widehat{K}^2$, and $\widehat{K}^2 \le \widehat{K}$. Consequently, $\mathrm{TV}(\widehat{K}(1-\widehat{K})) \le \mathrm{TV}(\widehat{K})+\mathrm{TV}(\widehat{K}^2) \le 2\mathrm{TV}(\widehat{K})$. Furthermore $\int |\widehat{b}_K(\omega)\widehat{K}(\omega)|^2 d\omega \le \int |\widehat{b}_K(\omega)|^2 d\omega = \frac{1}{2\pi}\int b_K^2(x)dx$.
Let us now examine the first term in (30).

$$
\begin{aligned}
&\int \left(\widetilde{f}_K(x) - E\widetilde{f}_K(x)\right)^2 dx \\
=\; & \int \left(\frac{1}{n}\sum_{i=1}^{n} K(X_i - x) - (K * f)(x)\right)^2 dx \\
=\; & \frac{1}{n^2}\sum_{i,j}\int K(X_i - x)K(X_j - x)dx - \frac{2}{n}\sum_{i=1}^{n}\int K(X_i - x)(K * f)(x)dx + \int E^2[K(X-x)]dx \\
=\; & \frac{1}{n^2}\sum_{i,j}(K * K)(X_i - X_j) - \frac{1}{n}\sum_{i=1}^{n}\Big((K * K) * f\Big)(X_i) - \frac{1}{n}\sum_{j=1}^{n}\Big((K * K) * f\Big)(X_j) \\
& + \int E^2[K(X-x)]dx \\
=\; & \frac{1}{n^2}\sum_{i\neq j}\Big[(K * K)(X_i - X_j) - E[(K * K)(X_i - Y)|X_i] - E[(K * K)(X - X_j)|X_j]\Big] \\
& + \frac{1}{n}\int K^2(x)dx + \int E^2[K(X-x)]dx - \frac{2}{n^2}\sum_{i=1}^{n}\Big((K * K) * f\Big)(X_i) \qquad (33)
\end{aligned}
$$

Subtracting its expectation from (33), we obtain again a degenerate U-statistic, this time in $K * K$ instead of K, plus an inferior term.

$$
\begin{aligned}
&\int \left(\widetilde{f}_K(x) - E\widetilde{f}_K(x)\right)^2 dx - E\int \left(\widetilde{f}_K(x) - E\widetilde{f}_K(x)\right)^2 dx \\
=\; & \frac{1}{n^2}\sum_{i\neq j}\Big[(K * K)(X_i - X_j) - E[(K * K)(X_i - Y)|X_i] - E[(K * K)(X - X_j)|X_j]\Big]
\end{aligned}
$$

$$
\begin{aligned}
& -E\Big[(K*K)(X_i-X_j)-E[(K*K)(X_i-Y)|X_i]-E[(K*K)(X-X_j)|X_j]\Big] \\
& -\frac{2}{n^2}\Big[\sum_{i=1}^{n}\Big((K*K)*f\Big)(X_i)-E\Big((K*K)*f\Big)(X_i)\Big] \\
= \quad & \frac{n-1}{n}\Big[\frac{1}{n(n-1)}\sum_{i\neq j}(K*K)(X_i-X_j)-E[(K*K)(X_i-X_j)|X_i] \\
& -E[(K*K)(X_i-X_j)|X_j]+E\left[(K*K)(X_i-X_j)\right]\Big] \\
& -\frac{2}{n^2}\Big[\sum_{i=1}^{n}\Big((K*K)*f\Big)(X_i)-E\Big((K*K)*f\Big)(X_i)\Big] \\
=: \quad & \frac{n-1}{n}\Big[\frac{1}{n(n-1)}\sum_{i\neq j}U_{K*K}(X_i,X_j)\Big]-\frac{2}{n^2}\Big[\sum_{i=1}^{n}\Big((K*K)*f\Big)(X_i)-E\Big((K*K)*f\Big)(X_i)\Big] \quad (34)
\end{aligned}
$$

The inferior term could now be handled like the empirical process of $K*b_K$. But it already suffices to bound it by its maximum.

$$
\begin{aligned}
\frac{2}{n^2}\Big|\sum_{i=1}^{n}\Big((K*K)*f\Big)(X_i)-E\Big((K*K)*f\Big)(X_i)\Big| \quad & \leq \quad \frac{4}{n}\max|(K*K)*f| \\
& \leq \quad \frac{4}{n}\sqrt{\int K^2(x)dx}\;\|f\|_2
\end{aligned}
$$

U_{K*K} is decomposed analogously to U_K in Subsection 4.1. Define:

$$
\alpha_{st}(K*K) \;:=\; \int \widehat{K}^2(\omega)\widehat{\varphi}_{st}(\omega)d\omega \qquad \text{and} \qquad \beta_{st}(K*K) \;:=\; \int \widehat{K}^2(\omega)\widehat{\psi}_{st}(\omega)d\omega
$$

The set of "favorable events" is the same as for U_K, A_{n1}, on which it holds that (see Lemma I.4.1):

$$
\begin{aligned}
\Big|\frac{1}{n(n-1)}\sum_{i\neq j}U_{K*K}(X_i,X_j)\Big| \quad = \quad & \Big|\alpha_{01}(K*K)\left[\frac{1}{n(n-1)}\sum_{i\neq j}U_{\varphi_{01}}(X_i,X_j)\right] \\
& +\sum_{s=0}^{-d_n}\alpha_{s2}(K*K)\left[\frac{1}{n(n-1)}\sum_{i\neq j}U_{\varphi_{s2}}(X_i,X_j)\right] \\
& +\sum_{s=-1}^{-d_n}\sum_{t=2}^{2^s n}\beta_{-st}(K*K)\left[\frac{1}{n(n-1)}\sum_{i\neq j}U_{\psi_{-st}}(X_i,X_j)\right] \\
& +\sum_{s=0}^{\infty}\sum_{t=1}^{2^s n}\beta_{st}(K*K)\left[\frac{1}{n(n-1)}\sum_{i\neq j}U_{\psi_{st}}(X_i,X_j)\right]\Big| \\
= \quad & O\left(\frac{\ln^{5/2}n}{n}\right)\left(\sqrt{\int \widehat{K}^4(\omega)d\omega}+\mathrm{TV}(\widehat{K}^2)\right) \\
= \quad & O\left(\frac{\ln^{5/2}n}{n}\right)\left(\sqrt{\int K^2(x)dx}+1\right) \qquad (35)
\end{aligned}
$$

Hence, on A_n we obtain:

$$
\begin{aligned}
& |ISE(K) - MISE(K)| \\
\leq\ & \frac{n-1}{n}\Big|\frac{1}{n(n-1)}\sum_{i\neq j} U_{K*K}(X_i, X_j)\Big| + \frac{2}{n}\Big|\sum_{j=1}^{n}(b_K * K)(X_j) - E[(b_K * K)](X_j)\Big| \\
& - \frac{2}{n^2}\Big[\sum_{i=1}^{n}\Big((K*K)*f\Big)(X_i) - E\Big((K*K)*f\Big)(X_i)\Big] \\
=\ & O\left(\frac{\ln^{5/2} n}{n}\right)\left(\sqrt{\int K^2(x)dx} + 1\right) + O\left(\frac{\ln^2 n}{\sqrt{n}}\right)\left(\sqrt{\int b_K^2(x)dx}\ +\ \frac{\|f\|_2}{\sqrt{n}}\right) \\
& + O(n^{-1})\sqrt{\int K^2(x)dx}\ \|f\|_2 \\
=\ & O(n^{-\delta})MISE(K) + O\Big(n^{\delta-1}\Big) \qquad \square\ \mathbf{A2}
\end{aligned}
$$

Therewith, all propositions are verified and the proof of Theorem I.2.1 is valid.

6 PRACTICAL COMPUTATION

Once the statistical properties of the CV-optimal kernel function K_0 have been examined, we would like to actually compute this kernel from a sample $X_1, \ldots, X_n$. K_0 is $\arg\min CV(K)$ within the set $\mathcal{K}_n := \{K \in \mathcal{K} | \text{supp}\, \widehat{K} \subseteq (-n, n)\}$. Hence we face a minimization problem. As already noted in Section 1, the set $\mathcal{K}$ is convex. With respect to the properties $\widehat{K}(\omega) \in \mathbb{R}$ and $\widehat{K}(\omega) \geq 0$, the convexity is clear. For the L_2-norm we have:

$$
\begin{aligned}
\int\Big(\lambda K_1(x) + (1-\lambda)K_2(x)\Big)^2 - \lambda K_1^2(x) - (1-\lambda)K_2^2(x)dx \ &=\ \lambda(\lambda-1)\int\Big(K_1(x) - K_2(x)\Big)^2 dx \\
&\leq\ 0
\end{aligned}
$$

Given that all $\widehat{K}$ in $\mathcal{K}$ are unimodal and symmetric around 0, their mode is 0. And a convex combination of any two $\widehat{K}$ is again unimodal. Convexity is also preserved through the trimming of the support of $\widehat{K}$. And furthermore, $CV(K)$ is a convex function:

$$
\begin{aligned}
& CV\Big(\lambda K_1 + (1-\lambda)K_2\Big) - \lambda CV(K_1) - (1-\lambda)CV(K_2) \\
=\ & \int\Big(\lambda \widetilde{f}_{K_1}(x) + (1-\lambda)\widetilde{f}_{K_2}(x)\Big)^2 dx - \frac{2}{n(n-1)}\sum_{j\neq l}\Big[\lambda K_1(X_j - X_l) + (1-\lambda)K_2(X_j - X_l)\Big] \\
& - \lambda\left[\int \widetilde{f}_{K_1}(x)^2 dx - \frac{2}{n(n-1)}\sum_{j\neq l} K_1(X_j - X_l)\right] \\
& - (1-\lambda)\left[\int \widetilde{f}_{K_2}(x)^2 dx - \frac{2}{n(n-1)}\sum_{j\neq l} K_2(X_j - X_l)\right] \\
=\ & \int\Big(\lambda \widetilde{f}_{K_1}(x) + (1-\lambda)\widetilde{f}_{K_2}(x)\Big)^2 - \lambda \widetilde{f}_{K_1}(x)^2 - (1-\lambda)\widetilde{f}_{K_1}(x)^2 dx \\
=\ & \lambda(\lambda-1)\int\Big(\widetilde{f}_{K_1}(x) - \widetilde{f}_{K_2}(x)\Big)^2 dx \\
\leq\ & 0
\end{aligned}
$$

and strictly < 0, for λ different from 0 and 1, i.e. CV is even strictly convex. Therefore $\min CV(K)$ over $\mathcal{K}_n$ is a convex optimization problem, with an argument that is itself a non-increasing function, $\widehat{K}$: $[0,n) \longrightarrow [0,1]$.
For a discrete version of $\mathcal{K}_n$, say $\mathcal{K}_n^t$, which contains all real, symmetric and unimodal piece-wise constant functions on $[0,n)$, with jumps at the points $2^{-t}k$, $k = 1 \ldots 2^t n$, and values $\in [0,1]$,

$$\begin{aligned}
\widehat{K}^t(\omega) &= \sum_{k=1}^{2^t n} a_k I\left(2^{-t}(k-1) \le |\omega| < 2^{-t}k\right), \qquad \text{where } 1 \ge a_1 \ge \ldots \ge a_{2^t n} \ge 0 \\
K^t(x) &= \frac{1}{\pi x} \sum_{k=1}^{2^t n} a_k \Big(\sin(2^{-t}kx) - \sin(2^{-t}(k-1)x)\Big) \\
\int K^t(x)^2 dx &= \frac{1}{\pi}\, 2^{-t} \sum_{k=1}^{2^t n} a_k^2 \;\le\; n
\end{aligned}$$

the minimization of $CV(K)$ over $\mathcal{K}_n^t$ is still a convex optimization problem, but now with respect to a parameter of dimension $2^t n$ (number of variables) and with $2^t n + 2$ constraints. The L_1-distance between a kernel function $\widehat{K}$ in $\mathcal{K}_n$ and its closest neighbor $\widehat{K}^t$ in $\mathcal{K}_n^t$ is not greater than 2^{-t} (and thus the same applies for the supremum distance between K and K^t). It follows that

$$\begin{aligned}
& \big|CV(K) - CV(K^t)\big| \\
=\; & \Big|\frac{1}{n}\int K(x)^2 - K^t(x)^2 dx + \frac{1}{n^2}\sum_{j\ne l}\Big[K * K(X_j - X_l) - K^t * K^t(X_j - X_l)\Big] \\
& - \frac{2}{n(n-1)}\sum_{j\ne l}\Big[K(X_j - X_l) - K^t(X_j - X_l)\Big]\Big| \\
=\; & \frac{1}{2\pi}\Big|\frac{1}{n}\int \widehat{K}(\omega)^2 - \widehat{K}^t(\omega)^2 d\omega + \frac{1}{n^2}\sum_{j\ne l}\int \Big(\widehat{K}(\omega)^2 - \widehat{K}^t(\omega)^2\Big)\, e^{-i\omega(X_j - X_l)} d\omega \\
& - \frac{2}{n(n-1)}\sum_{j\ne l}\int \Big(\widehat{K}(\omega) - \widehat{K}^t(\omega)\Big)\, e^{-i\omega(X_j - X_l)} d\omega\Big| \\
\le\; & \frac{1}{2\pi}\left(\frac{1}{n}\int 2|\widehat{K}(\omega) - \widehat{K}^t(\omega)| d\omega + \frac{1}{n^2}\sum_{j\ne l}\int 2|\widehat{K}(\omega) - \widehat{K}^t(\omega)| d\omega \right. \\
& \left. + \frac{2}{n(n-1)}\sum_{j\ne l}\int |\widehat{K}(\omega) - \widehat{K}^t(\omega)| d\omega\right) \\
\le\; & \frac{2}{\pi}\int |\widehat{K}(\omega) - \widehat{K}^t(\omega)| d\omega \\
\le\; & \frac{2}{\pi}\cdot 2^{-t}
\end{aligned}$$

$$\Longrightarrow\; CV(K_0^t) \;:=\; \min_{\mathcal{K}_n^t} CV(K) \;\le\; CV\left((K_0)^t\right) \;\le\; CV(K_0) + \frac{2}{\pi}\cdot 2^{-t}$$

$\mathcal{K}_n^t \subseteq \mathcal{K}_n^{t+1}$, so that the sequence $CV(K_0^t)$, in $t \longrightarrow \infty$, decreases monotonously towards $CV(K_0)$.

Since CV is strictly convex, for any t, the minimization problem $\min\{CV(K)|K \in \mathcal{K}_n^t\}$ possesses a unique solution K_0^t, and the sequence $\{K_0^t\}_{t\in\mathbb{N}}$ converges towards K_0, the unique solution of the original problem.

It would be an interesting question to consider if a profound analysis in terms of optimization could yield a more sophisticated solution to the problem, possibly avoiding discretization and giving convergence rates over classes of densities.

6.1 Simulation

By a simulation study we try to get an impression of how the new method for selecting the kernel function works in practice. For this purpose, several density functions f have been chosen in order to generate sample data at different sample sizes n. A computer program calculates the CV-optimal (piecewise constant) kernel K_0 for each sample, its corresponding $ISE(K_0)$ and, repeating the exercise $b = 50$ times, its mean value.
For the sake of comparison with conventional kernel estimates, we have additionally computed ISE for fixed kernel functions with CV-optimal bandwidth, namely the second order Gauss kernel and the rate adaptive (but not minimax) sinc-kernel.
The following density functions will be considered in the simulation: The Standard Normal distribution density with exponentially decreasing Fourier transform and $f(x) = \frac{1-\cos 5x}{5\pi x^2}$, an entire density with $\widehat{f}(\omega) = (1 - |\omega|/5)_+$, whose both particular smoothness is not exploited by most of the common kernel functions.
On the other hand two density functions with slowly decreasing Fourier transform: $\frac{1}{c}(1 + 0.3\sin x + 0.2\cos 3x)(1 - (x/\pi)^{10})_+$, trimodal and with bounded support, and the density of a Γ-distribution with parameters 2 and 3/2, which is skewed and only defined on $\mathbb{R}^+$.
The simulation was conducted employing the commercial statistical software package S-PLUS, along with the optimization module NUOPT. We recognize that even the roughly cross-validated kernel functions (the program is documented in Appendix B) exhibit a good performance, the gain is especially pronounced for densities with well behaved Fourier transforms, and undeniable for those which are more difficult to estimate.

$$f(x) = \frac{1}{2\pi}\, e^{-x^2/2}$$

n	CV optimal K_0	Gaussian $\exp\{-x^2/2\}$	sinc $x^{-1}\sin x$
50	0.0118	0.0172	0.0216
100	0.0065	0.0127	0.0114
500	0.0013	0.0028	0.0026
1000	0.0008	0.0016	0.0010

$$f(x) = \frac{1-\cos 5x}{5\pi x^2}$$

n	CV optimal K_0	Gaussian $\exp\{-x^2/2\}$	sinc $x^{-1}\sin x$
50	0.0229	0.0307	0.0392
100	0.0079	0.0184	0.0187
500	0.0022	0.0055	0.0055
1000	0.0013	0.0022	0.0023

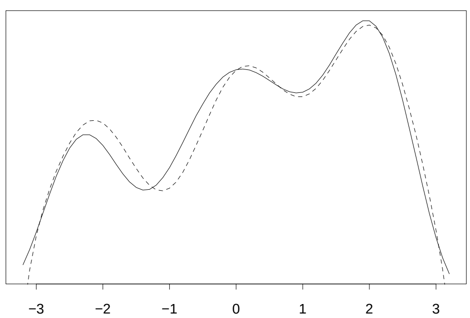
$$f(x) \;=\; \tfrac{1}{c}\;\left(1+\tfrac{3\sin x}{10}+\tfrac{\cos 3x}{5}\right)\left(1-\tfrac{x^{10}}{\pi^{10}}\right)_+$$

n	CV optimal K_0	Gaussian $\exp\{-x^2/2\}$	sinc $x^{-1}\sin x$
50	0.0110	0.0205	0.0249
100	0.0073	0.0125	0.0145
500	0.0025	0.0040	0.0047
1000	0.0018	0.0021	0.0025

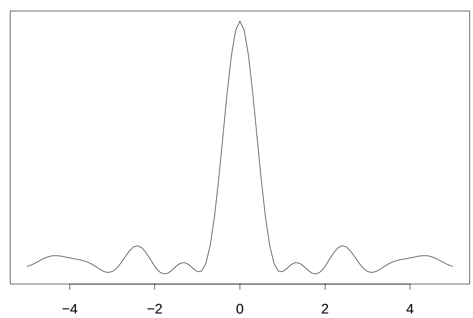
$$f(x) \;=\; \tfrac{(3/2)^2}{\Gamma(2)}\;xe^{-3x/2}$$

n	CV optimal K_0	Gaussian $\exp\{-x^2/2\}$	sinc $x^{-1}\sin x$
50	0.0270	0.0277	0.0617
100	0.0148	0.0209	0.0403
500	0.0052	0.0061	0.0166
1000	0.0027	0.0031	0.0037

The plots below for $f(x) = \frac{1}{c}\,(1+0.3\sin x+0.2\cos 3x)\,(1-(x/\pi)^{10})_+$ at $n=500$, show that the procedure delivers nice and smooth estimates, avoiding any absurd features, which are from time to time observed in certain estimation techniques. The kernel in contrast, is a bit unorthodox.



density and estimator



kernel function

If a conclusion should be drawn from this fairly short study, the implementation of kernel-CV seems to be worth an effort, since a quite important improvement in estimation power is already attainable on relatively little expenses.

To tap the full potential of the cross-validation kernel choice, it would be highly desirable to develop an elaborate numerical algorithm for the minimization of CV not only over step functions but over all unimodal functions with support in $(-n, n)$.

Part II
NON-PARAMETRIC REGRESSION

1 PROBLEM, THEOREM and PROOF

In the second part of the present work we attempt to adapt our cross validation technique for the choice of the optimal kernel to the problem of estimating a regression function.

$$Y = m(x) + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0,1) \tag{1}$$

Let us assume that we have a design with equidistant points of observation on $[0,1]$: $x_1, \ldots, x_n$, $x_i = \frac{2i-1}{2n}$, and corresponding observations $Y_1, \ldots, Y_n$ with $Y_i = m(x_i) + \varepsilon_i$, ε_i independent and identically standard-normally distibuted. We apply the Gasser-Müller estimator with kernel function K:

$$\widetilde{m}_K(x) \;=\; \sum_{i=1}^{n} k_i(x) Y_i \;=\; \sum_{i=1}^{n} \int_{\frac{x_{i-1}+x_i}{2}}^{\frac{x_i+x_{i+1}}{2}} K(z-x)dz\, Y_i \tag{2}$$

k_1 and k_n computed with $x_0 := -\frac{1}{2n}$und $x_{n+1} := \frac{2n+1}{2n}$. Let the regression function m be periodic with period length 1. (The reason for such an artificial assumption is no other than to avoid boundary effects in the estimator. Evidently, it is possible to drop it, but not without a loss in readability.) If we want to carry this property forward to $\widetilde{m}_K$, the kernel function K also has to be periodic with period 1. So when estimating m at a point x near to 0, observations at points near to 1 are taken into account with weights as high as if they had been observed at points to the left side of 0. Vice versa, when estimating m at points x near to 1.
Two further points, $x_{-1} := -\frac{3}{2n}$ and $x_{n+2} := \frac{2n+3}{2n}$ are necessary to define the cross-validation (leave-one-out) criterion:

$$CV(K) \;=\; \frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{m}_K^{(-i)}(x_i) - Y_i\right)^2 \;=\; \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j\neq i} k_j^{(-i)}(x_i)Y_j - Y_i\right)^2 \tag{3}$$

$$\text{where} \qquad k_{i-1}^{(-i)}(x_i) = \int_{\frac{x_{i-2}+x_{i-1}}{2}}^{x_i} K(z-x)dz, \quad k_{i+1}^{(-i)}(x_i) = \int_{x_i}^{\frac{x_{i+1}+x_{i+2}}{2}} K(z-x)dz$$

$$\text{and} \qquad k_j^{(-i)}(x_i) = k_j(x_i), \quad \text{otherwise}$$

As n tends to ∞, the kernel function is customarily rescaled by a bandwidth $h = h_n \longrightarrow 0$, $nh \longrightarrow \infty$: $K_h(.) = h^{-1}K(./h)$, and statements about the choice of bandwidth and kernel function similar to the case of density estimation can be found in: Hall (1984) – bandwidth CV; Gasser, Müller, Mammitzsch (1985) – pseudo-optimal kernel functions; Golubev (1987) – minimax kernel functions; Kneip (1994) simultaneous choice of kernel order and bandwidth. Kneip (1994) chooses however the order of the kernel function only within a finite subset of $\mathbb{N}$. Super kernels are again rate-adaptive but not asymptotically minimax (with respect to the constant).
In order to approximate the MISE-optimal kernel function in a more flexible way corresponding to the procedure for density estimation, we will allow for kernel functions in the

set $\mathcal{K}'$ of all bounded 1-periodic functions with $\int_0^1 K(x)dx = 1$ and real, non-negative, unimodal Fourier expansion $\widehat{K}(2\pi\kappa) := \int_0^1 K(x)e^{i2\pi\kappa x}dx$ and $\|K\|_2 := \int_0^1 K^2(x)dx \leq \sqrt{n}$ (see Appendix A.1 for a discussion of the exponential Fourier expansion).
Let m be bounded, its first and its second derivative exist and be bounded, too. Again let K^* be the MISE-optimal kernel out of $\mathcal{K}'$. Denote the CV-optimal kernel out of $\mathcal{K}'_n := \{K \in \mathcal{K}' | \widehat{K}(2\pi\kappa) = 0 \text{ for } |\kappa| > \sqrt{n}\}$ by K_0.

THEOREM II.1.1 Under the hypotheses on m, ε and K stated in this section, it holds for any $\delta > 0$:

$$|E[ISE(K_0)] - MISE(K^*)| = O(n^{-\delta})MISE(K^*) + O\left(n^{\delta-1}\right)$$

Remark 1 For uniformly bounded $|m^{(j)}|$, $j = 0, 1, 2$, also $\|m\|_2$ is uniformly bounded, and the result of Theorem II.1.1 is uniform over regression functions m. The restrictions on $|m'|$ and $|m''|$ are due to the discrete nature of CV used to approximate MISE (uniform boundedness of the derivatives of m is but one means, yet uniform convergence of E[CV] towards MISE has to be ensured).

Remark 2 In comparison to the density case, we had to impose stronger restrictions to the class of kernel functions available to the minimization of CV: $\widehat{K}(2\pi\kappa) = 0$ for $|\kappa| > \sqrt{n}$. But since uniformity over m is anyway only achieved for $|m'|$ and $|m''|$ uniformly bounded, by the same assumption we derive $MISE(K^*_{(\sqrt{n})}) = MISE(K^*) + O(n^{-1})$ with uniform $O(.)$.

Remark 3 As already noted earlier, the somewhat exclusive design of the estimation problem (periodic regression function, but also i.i.d. standard-normal errors) is possible to overcome. We have chosen it with the intention not to further complicate the calculations, which are anyway quite technical.

PROOF of THEOREM II.1.1 The proof runs along the lines of Section 3 Part I on density estimation. The following propositions will be proven in Sections 2, 3 and 4: For any $\lambda < \infty$ there exists a set of "favorable events" $B_n \subseteq \mathbb{R}^n$, such that for an abitrary observation $(Y_1, \ldots, Y_n) = (m(x_1) + \varepsilon_1, \cdots, m(x_n) + \varepsilon_n)$ with $(\varepsilon_1, \ldots, \varepsilon_n) = \varepsilon \in B_n$ and for any $\delta > 0$, it holds that

$$\begin{aligned}
\textbf{B1} \quad |ISE(K) - CV(K)| &\leq |ISE(K) - MISE(K)| + |MISE(K) - E[CV(K)]| \\
&\quad + |E[CV(K)] - CV(K)| \\
&= O(n^{-\delta})MISE(K) + O\left(n^{\delta-1}\right) \\
&\quad \forall\, K \in \left\{K \in \mathcal{K}' | \widehat{K}(2\pi\kappa) = 0 \text{ for } |\kappa| > \sqrt{n}\right\} \\
\textbf{B2} \quad P\left(\varepsilon \in B_n^c\right) &= O\left(n^{-\lambda}\right)
\end{aligned}$$

where all $O(.)$ are uniform under the assumption. When λ is chosen sufficiently large, these relations and the fact that $E[ISE^2(K)] = O(n^2)$, which will be proven below, enable us to show:

$$\begin{aligned}
E[ISE(K_0)] - MISE(K^*) &= E\left[ISE(K_0) - ISE(K^*)\right] \\
&\leq E_{B_n}\left[ISE(K_0) - ISE(K^*)\right] + E_{B_n^c}\left[ISE(K_0)\right] \\
&\leq E_{B_n}\left[ISE(K_0) - CV(K_0)\right] + 0 + E_{B_n}\left[CV(K^*) - ISE(K^*)\right]
\end{aligned}$$

$$
\begin{aligned}
& +\sqrt{P(B_n^c)E[ISE^2(K_0)]} \\
= \;& O(n^{-\delta})E[ISE(K_0)] + O(n^{-\delta})MISE(K^*) + O\Big(n^{\delta-1}\Big) \\
& + O\left(n^{-\lambda/2}\right) O\left(n\right)
\end{aligned}
$$

For the opposite difference, it holds:

$$
MISE(K^*) - E[ISE(K_0)] = O(n^{-\delta})MISE(K^*) + O\Big(n^{\delta-1}\Big)
$$

And therefore:

$$
|E[ISE(K_0)] - MISE(K^*)| = O(n^{-\delta})MISE(K^*) + O\Big(n^{\delta-1}\Big)
$$

The missing gap for the completion of the proof is closed by:

$$
\begin{aligned}
E[ISE^2(K)] &= E\left[\int_0^1 \Big(\widetilde{m}_K(x) - m(x)\Big)^2 dx\right]^2 \\
&\leq 4E\left[\int_0^1 \widetilde{m}_K^2(x)dx + \int_0^1 m(x)^2 dx\right]^2 \\
&\leq \text{const. } \|K\|_2^4 \left(\max|m|^2 + 2\max|m|E\varepsilon_i^2 + n^{-1}E\varepsilon_i^4 + (n-1)n^{-1}E^2\varepsilon_i^2\right) \\
&\quad + \text{const. } \|m\|_2^4 \\
&= O(n^2)
\end{aligned}
$$

where the constants do neither depend on m nor on K. $\square$

2 CV-E[CV]

Let us introduce a matrix notation for the Gasser-Müller weights:

$$
\begin{aligned}
K^{\pm}(x) &:= \int_{-1/(2n)}^{1/(2n)} K\left(x - z\right) dz \\
\mathbb{K} &:= \Big(\Big(K^{\pm}(x_i - x_j)\Big)\Big)_{i,j=1}^n \qquad \text{such that} \qquad \Big(\widetilde{m}_K(x_i)\Big)_{i=1}^n = \mathbb{K}Y
\end{aligned}
\tag{4}
$$

In the computation of the cross-validation criterion, the weight matrix changes: Let $\mathbb{C}$ be an $n \times n$ matrix, with all elements in the diagonal equal to -1, all elements in the secondary diagonals equal to $1/2$ and 0 elsewhere, except for $(\mathbb{C})_{1n} = (\mathbb{C})_{n1}$, which be again set to $1/2$.

$$
CV(K) = \frac{1}{n}\Big(\left(\mathbb{K} + K^{\pm}(0)\mathbb{C}\right)Y - Y\Big)^T \Big(\left(\mathbb{K} + K^{\pm}(0)\mathbb{C}\right)Y - Y\Big) \tag{5}
$$

With definition

$$
\begin{aligned}
b_{\mathbb{K}}(x) &:= E[\widetilde{m}_K(x)] - m(x) = \sum_{i=1}^n K^{\pm}(x - x_i)m(x_i) - m(x) \\
\ell_K(x) &:= (K * m)(x) - m(x)
\end{aligned}
\tag{6}
$$

we shall start proving the third part of proposition **B1** by a consideration of MISE(K), so as to figure out the appropriate size of the bound for CV-E[CV].

$$
\begin{aligned}
|b_{\mathbb{K}}(x) - \mathscr{b}_K(x)| &= \Big|\sum_{i=1}^{n}\int_{-1/(2n)}^{1/(2n)} K(x-x_i-y)dy\; m(x_i) - \int_0^1 K(x-y)m(y)dy\Big| \\
&= \Big|\sum_{i=1}^{n}\int_{-1/(2n)}^{1/(2n)} K(x-x_i-y)m(x_i) - K(x-x_i-y)m(x_i+y)dy\Big| \\
&\le \sum_{i=1}^{n}\int_{-1/(2n)}^{1/(2n)} |K(x-x_i-y)|dy\; \frac{1}{n}\max|m'| \\
&\le \frac{1}{n}\max|m'|\sqrt{\int_0^1 K^2(x)dx} \qquad (7)
\end{aligned}
$$

For the next approximation we need that from $\widehat{K}(2\pi\kappa)=0$ or $|\kappa|>\sqrt{n}$ it follows that

$$
\begin{aligned}
\max|K''| &\le \sum_{|\kappa|\le\sqrt{n}} (2\pi\kappa)^2\widehat{K}(2\pi\kappa) \\
&\le \sum_{|\kappa|\le\sqrt{n}} (2\pi\kappa)^2 \\
&\le (2\pi)^2\cdot 2n^{3/2}
\end{aligned}
$$

and as well $\int K'(x)^2dx = \sum(2\pi\kappa)^2\widehat{K}^2(2\pi\kappa) \le (2\pi)^2\cdot 2n^{3/2}$.

$$
\begin{aligned}
|K^{\pm}(x) - \frac{1}{n}\,K(x)| &= \Big|\int_{-1/(2n)}^{1/(2n)} K(x+y)dy - \frac{1}{n}\,K(x)\Big| \\
&= \Big|\int_{-1/(2n)}^{1/(2n)} K(x+y) - K(x)dy\Big| \\
&\le \frac{1}{n^2}|K'(x)| + O(n^{-3/2}) \qquad (8)
\end{aligned}
$$

and therefore

$$
\begin{aligned}
&\Big|MISE(K) - \int_0^1 \mathscr{b}_K^2(x)dx - \frac{\sigma^2}{n}\int_0^1 K^2(x)dx\Big| \\
=\;& \Big|E\int_0^1\Big(\widetilde{m}_K(x) - m(x)\Big)^2dx - \int_0^1 \mathscr{b}_K^2(x)dx - \frac{\sigma^2}{n}\int_0^1 K^2(x)dx\Big| \\
=\;& \Big|\int_0^1 b_{\mathbb{K}}^2(x)dx - \int_0^1 \mathscr{b}_K^2(x)dx + \sigma^2\sum_{k=1}^{n}\int_0^1 K^{\pm}(x_k-x)^2dx - \frac{\sigma^2}{n}\int_0^1 K^2(x)dx\Big| \\
=\;& \Big|\int_0^1 b_{\mathbb{K}}^2(x)dx - \int_0^1 \mathscr{b}_K^2(x)dx + \sigma^2\sum_{k=1}^{n}\int_0^1 K^{\pm}(x_k-x)^2 - \frac{1}{n}\,K^2(x_k-x)dx\Big| \\
=\;& \int_0^1\Big(b_{\mathbb{K}}(x) - \mathscr{b}_K(x)\Big)\Big(2\mathscr{b}_K(x) + b_{\mathbb{K}}(x) - \mathscr{b}_K(x)\Big)dx \\
&+\sigma^2\sum_{k=1}^{n}\int_0^1\left(K^{\pm}(x_k-x) - \frac{1}{n}\,K(x_k-x)\right)\left(\frac{2}{n}\,K(x_k-x) + K^{\pm}(x_k-x) - \frac{1}{n}\,K(x_k-x)\right)dx
\end{aligned}
$$

$$\leq \;\; O(n^{-1})\max|m'|\sqrt{\int_0^1 K^2(x)dx}\sqrt{\int_0^1 b_K^2(x)dx} + O(n^{-2})\max|m'|^2\int_0^1 K^2(x)dx$$
$$+O(n^{-5/4})\sigma^2\sqrt{\int_0^1 K^2(x)dx} + \sigma^2 O(n^{-7/4})$$

which means that

$$MISE(K) \;\; = \;\; \left(\int_0^1 b_K^2(x)dx + \frac{\sigma^2}{n}\int_0^1 K^2(x)dx\right)\left(1+O(n^{-1/2})\right) \tag{9}$$

As residual terms containing functionals of the regression function m will from now on appear oftentimes, we can, for the sake of readability, not preserve a notation, where O's are independent of m. The uniformity of the results is nevertheless guaranteed for uniformly bounded $|m^{(j)}|$, $j = 0, 1, 2$.
Using matrix notation for CV and

$$m \;:=\; \Big(m(x_i)\Big)_{i=1}^n \qquad \varepsilon \;:=\; \Big(\varepsilon_i\Big)_{i=1}^n \qquad b_{\mathbb{K}} \;:=\; \Big(b_{\mathbb{K}}(x_i)\Big)_{i=1}^n \tag{10}$$

$$\begin{aligned} CV(K) \;\; &= \;\; \frac{1}{n}\Big(\left(\mathbb{K}+K^{\pm}(0)\mathbb{C}\right)Y - Y\Big)^T\Big(\left(\mathbb{K}+K^{\pm}(0)\mathbb{C}\right)Y - Y\Big) \\ &= \;\; \frac{1}{n}Y^T\Big(\mathbb{K}+K^{\pm}(0)\mathbb{C}-I\Big)^2 Y \\ &= \;\; \frac{1}{n}m^T\Big(\mathbb{K}+K^{\pm}(0)\mathbb{C}-I\Big)^2 m + \frac{2}{n}b_{\mathbb{K}}^T\Big(\mathbb{K}+2K^{\pm}(0)\mathbb{C}-I\Big)\varepsilon \\ &\quad\; +\frac{2}{n}m^T K^{\pm}(0)^2\mathbb{C}^2\varepsilon + \frac{1}{n}\varepsilon^T\Big(\mathbb{K}^2+2K^{\pm}(0)\mathbb{K}\mathbb{C}-2\mathbb{K}\Big)\varepsilon \\ &\quad\; +\frac{1}{n}\varepsilon^T\Big(K^{\pm}(0)^2\mathbb{C}^2-2K^{\pm}(0)\mathbb{C}+I\Big)\varepsilon \end{aligned} \tag{11}$$

The term $\varepsilon^T\varepsilon$ is independent from K and thus drops out of $ISE(K_0) - CV(K_0) + CV(K^*) - ISE(K^*)$. Since we have assumed that $\widehat{K}(2\pi\kappa) = 0$ for $n > \sqrt{n}$, the high-frequency component of the bias does not depend on K,

$$h_m(x) \;\; := \;\; \sum_{|\kappa|>\sqrt{n}} \widehat{m}(2\pi\kappa)e^{-i2\pi\kappa x} \;\; = \;\; -\sum_{|\kappa|>\sqrt{n}} \widehat{b}_{\mathbb{K}}(2\pi\kappa)e^{-i2\pi\kappa x} \tag{12}$$

and cancels out between $b_{\mathbb{K}_0} - b_{\mathbb{K}^*} = (b_{\mathbb{K}_0} + h_m) - (b_{\mathbb{K}^*} + h_m)$, where $h_m := \Big(h_m(x_i)\Big)_{i=1}^n$.
$m^T(\mathbb{K}+K^{\pm}(0)\mathbb{C}-I)^2 m$ is deterministic and vanishes after subtraction of $E[CV(K)]$.
$\varepsilon^T(K^{\pm}(0)^2\mathbb{C}^2 - 2K^{\pm}(0)\mathbb{C})\varepsilon$ and $m^T(K^{\pm}(0)^2\mathbb{C}^2 - K^{\pm}(0)\mathbb{C})\varepsilon$ only depend on K through $K^{\pm}(0)$, and we know that $K^{\pm}(0) = \int_{-1/(2n)}^{1/(2n)} K(x)dx \leq \frac{1}{n}\max|K| \leq \frac{1}{n}\sum|\widehat{K}(2\pi\kappa)| \leq \frac{2\sqrt{n}+1}{n}$. Hence we are left to examine

$$\begin{aligned} &\frac{2}{n}(b_{\mathbb{K}}+h_m)^T\Big(\frac{2(2\sqrt{n}+1)}{n}\mathbb{C}-I\Big)\varepsilon + \frac{2}{n}b_{\mathbb{K}}^T\mathbb{K}\varepsilon + \frac{8}{n^2}m^T\mathbb{C}^2\varepsilon + \frac{1}{n}\varepsilon^T\Big(\mathbb{K}^2+\frac{2(2\sqrt{n}+1)}{n}\mathbb{K}\mathbb{C}-2\mathbb{K}\Big)\varepsilon \\ &+\frac{1}{n}\varepsilon^T\Big(\frac{(2\sqrt{n}+1)^2}{n^2}\mathbb{C}^2+\frac{2(2\sqrt{n}+1)}{n}\mathbb{C}\Big)\varepsilon \end{aligned} \tag{13}$$

In Section 2.2 the quadratic forms in $\mathbb{K}$ and ε, centered by subtracting their expected values, will be decomposed on a set of "favorable events" B_{n1} by means of a wavelet inspired function basis, so as to find the bound $O(n^{-1}\ln^2 n)\|K\|_2$, whereby the probability of B_{n1}^c will be

shown to be of the form $O(n^{-\lambda+1/2})$ for sufficiently large λ. The centered quadratic forms in $\mathbb{C}$ without $\mathbb{K}$ are dominated by $\lambda\sqrt{n\ln n}$ on the set B_{n2}. Similarly, $P(B_{n2}^c) \leq O(n^{-\lambda})$. The linear forms in $b_{\mathbb{K}}$ and ε, with expected value 0, will be shown to be bounded by $O(n^{-1/2}\ln^{3/2}n)\|\mathscr{b}_K\|_2$ on the set B_{n3}, $P(B_{n3}^c) \leq O(n^{-\lambda^2+1/2})$, through another decomposition in Section 2.3. Finally $m^T\mathbb{C}^2\varepsilon$ is smaller than $\lambda\sqrt{n\ln n}$, $P(B_{n4}^c) \leq O(n^{-\lambda^2})$, on the set B_{n4}.
Therefore, on the union of all sets B_{ni}^c we obtain $P(\bigcup B_{ni}^c) = P(B_n^c) = O(n^{-\lambda'})$ with suitable $\lambda' < \infty$, and on $\bigcap B_{ni} = B_n$:

$$
\begin{aligned}
&\left|CV(K) - \frac{1}{n}\varepsilon^T\varepsilon - E\left[CV(K) - \frac{1}{n}\varepsilon^T\varepsilon\right] + \frac{2}{n}h_m^T\Big(K^{\pm}(0)\mathbb{C} - I\Big)\varepsilon\right| \\
\leq\ & \left|\frac{2}{n}(b_{\mathbb{K}} + h_m)^T\varepsilon\right| + \left|\frac{1}{n}\varepsilon^T\Big(\mathbb{K}^2 - 2\mathbb{K}\Big)\varepsilon - E\left[\frac{1}{n}\varepsilon^T\Big(\mathbb{K}^2 - 2\mathbb{K}\Big)\varepsilon\right]\right| \\
&+\frac{2(2\sqrt{n}+1)}{n}\left(\left|\frac{2}{n}(b_{\mathbb{K}} + h_m)^T\mathbb{C}\varepsilon\right| + \left|\frac{1}{n}\varepsilon^T(\mathbb{K}+I)\mathbb{C}\varepsilon - E\Big[\frac{1}{n}\varepsilon^T(\mathbb{K}+I)\mathbb{C}\varepsilon\Big]\right|\right) \\
&+\frac{(2\sqrt{n}+1)^2}{n^2}\left(\frac{2}{n}m^T\mathbb{C}^2\varepsilon + \frac{1}{n}\varepsilon^T\mathbb{C}^2\varepsilon - E\Big[\frac{1}{n}\varepsilon^T\mathbb{C}^2\varepsilon\Big]\right) \\
=\ & O\left(\frac{\ln^{3/2}n}{\sqrt{n}}\right)\sqrt{\int_0^1 \mathscr{b}_K^2(x)dx} + O\left(\frac{\ln^2 n}{n}\right)\sqrt{\int_0^1 K^2(x)dx} \\
&+\frac{4\sqrt{n}+2}{n}\left(O\left(\frac{\ln^{3/2}n}{\sqrt{n}}\right)\sqrt{\int_0^1 \mathscr{b}_K^2(x)dx} + O\left(\frac{\ln^2 n}{n}\right)\sqrt{\int_0^1 K^2(x)dx} + O\left(\frac{\ln n}{\sqrt{n}}\right)\right) \\
&+\frac{(2\sqrt{n}+1)^2}{n^2}\,O\left(\frac{\ln^{1/2}n}{\sqrt{n}}\right) \\
=\ & O(n^{-\delta})MISE(K) + O\Big(n^{\delta-1}\Big)
\end{aligned}
$$

2.1 Fourier expansions

This technical subsection contains the calculation of the Fourier coefficients of the different functions stemming from CV-E[CV]. These will be needed in the wavelet decompositions of Subsections 2.2 and 2.3.
Because of periodicity of K, the matrix $\mathbb{K}$ is circulant (Lütkepohl (1996)), it exhibits in particular a Toeplitz structure. Furthermore it is symmetric – the value of $K^{\pm}(x_i - x_j)$ is known to depend only on the difference $|x_i - x_j|$. And $K(x)$ can be expanded into a Fourier series as (see also Appendix A.1):

$$\widehat{K}(2\pi\kappa) \ := \ \int_0^1 K(x)\,e^{i2\pi\kappa x}dx \qquad \text{and} \qquad K(x) \ = \ \sum_{\kappa\in\mathbb{Z}} \widehat{K}(2\pi\kappa)e^{-i2\pi\kappa x} \tag{14}$$

$\mathbb{K}^2$ is again circulant. In order to express the components of $\mathbb{K}^2$ as function of $|x_i - x_j|$, too, we define a discrete concatenation of two functions, similar to the convolution. For given n and $x_1, \ldots x_n$ as in Section 1 of this part and $f, g : [0,1] \longrightarrow \mathbb{R}$ let:

$$(f \diamond g)\,(x) := \frac{1}{n}\sum_{k=1}^{n} f(x - x_k)g(x_k) \tag{15}$$

Due to the periodicity of K we can write:

$$
\begin{aligned}
\left(\mathbb{K}^2\right)_{ij} &= \sum_{k=1}^{n} \int_{-1/2n}^{1/2n} K(x_i - x_k + y)dy \int_{-1/2n}^{1/2n} K(x_j - x_k + z)dz \\
&= \sum_{k=1}^{n} \int_{-1/2n}^{1/2n} K\Big((x_i - x_j) - x_k + y\Big)dy \int_{-1/2n}^{1/2n} K(x_k - z)dz \\
&= \int_{-1/2n}^{1/2n} \sum_{k=1}^{n} K\Big((x_i - x_j) - x_k + y\Big) \int_{-1/2n}^{1/2n} K(x_k - z)dzdy \\
&= \left(\sum_{k=1}^{n} K(\,.\, - x_k)K^{\pm}(x_k)\right)^{\pm} (x_i - x_j) \\
&= n\left(K \diamond K^{\pm}\right)^{\pm}(x_i - x_j) \qquad (16)
\end{aligned}
$$

$$
\begin{aligned}
n(\widehat{K \diamond K^{\pm}})(2\pi\kappa) &= \int_0^1 \sum_{k=1}^{n} K(x - x_k) \int_{-1/2n}^{1/2n} K(x_k - z)dz\ e^{i2\pi\kappa x}dx \\
&= \sum_{k=1}^{n} \int_0^1 K(x - x_k)\ e^{i2\pi\kappa(x - x_k)}dx\ e^{i2\pi\kappa x_k} \int_{-1/2n}^{1/2n} K(x_k - z)dz \\
&= \widehat{K}(2\pi\kappa) \sum_{k=1}^{n} \int_{-1/2n}^{1/2n} K(x_k - z)\left(e^{i2\pi\kappa(x_k - z)} + e^{i2\pi\kappa x_k} - e^{i2\pi\kappa(x_k - z)}\right)dz \\
&= \widehat{K}(2\pi\kappa)\left(\widehat{K}(2\pi\kappa) + \sum_{k=1}^{n} \int_{-1/2n}^{1/2n} K(x_k - z)\left(e^{i2\pi\kappa x_k} - e^{i2\pi\kappa(x_k - z)}\right)dz\right)
\end{aligned}
$$

$$
\begin{aligned}
\left|\widehat{K}^2(2\pi\kappa) - n(\widehat{K \diamond K^{\pm}})(2\pi\kappa)\right| &\le \widehat{K}(2\pi\kappa) \sum_{k=1}^{n} \int_{-1/2n}^{1/2n} |K(x_k - z)\ 2\pi\kappa z|\,dz \\
&\le \widehat{K}(2\pi\kappa) \frac{2\pi|\kappa|}{2n} \sqrt{\int_0^1 K^2(z)dz}
\end{aligned}
$$

Similarly to the density case, $\widehat{K}(2\pi\kappa) \le \|K\|_2 |2\kappa + 1|^{-1/2}$. Since $\widehat{K}(2\pi\kappa) = 0$ for $|\kappa| > \sqrt{n}$, it holds that:

$$
\left|\widehat{K}^2(2\pi\kappa) - n(\widehat{K \diamond K^{\pm}})(2\pi\kappa)\right| = \int_0^1 K^2(x)dx\ O\left(n^{-3/4}\right) \qquad (17)
$$

Next we examine

$$
\begin{aligned}
(\mathbb{KC})_{ij} &= \frac{1}{2}K^{\pm}(x_i - x_{j-1}) - K^{\pm}(x_i - x_j) + \frac{1}{2}K^{\pm}(x_i - x_{j+1}) \\
&= \frac{1}{2}\int_{-1/(2n)}^{1/(2n)} K\left(x_i - x_{j-1} + z\right)dz - \int_{-1/(2n)}^{1/(2n)} K\left(x_i - x_j + z\right)dz \\
&\quad + \frac{1}{2}\int_{-1/(2n)}^{1/(2n)} K\left(x_i - x_{j+1} + z\right)dz \\
&= \frac{1}{2}\int_{1/(2n)}^{3/(2n)} K\left(x_i - x_j + z\right)dz - \int_{-1/(2n)}^{1/(2n)} K\left(x_i - x_j + z\right)dz \\
&\quad + \frac{1}{2}\int_{-3/(2n)}^{-1/(2n)} K\left(x_i - x_j + z\right)dz \qquad (18)
\end{aligned}
$$

It is obvious that the components of $\mathbb{KC}$ are constructed in exactly the same way as those of $\mathbb{K}$. The integrals are just taken over different intervals. Let us consider the bias. We remember $b_{\mathbb{K}}$ and define b_K:

$$\begin{aligned}
b_{\mathbb{K}}(x) &= n\left(K^{\pm}\diamond m\right)(x) - m(x) = \sum_{k=1}^{n} K^{\pm}(x-x_k)m(x_k) - m(x) \\
b_K(x) &:= \left(K\diamond m\right)(x) - m(x) = \frac{1}{n}\sum_{k=1}^{n} K(x-x_k)m(x_k) - m(x) \qquad (19)
\end{aligned}$$

$$\begin{aligned}
\frac{1}{n}\, b_{\mathbb{K}}(x) &= \frac{1}{n}\sum_{k=1}^{n} K^{\pm}(x-x_k)m(x_k) - m(x) \\
&= \int_{-1/2n}^{1/2n} \frac{1}{n}\sum_{k=1}^{n} K(x+y-x_k)m(x_k)dy - m(x) \\
&= \int_{-1/2n}^{1/2n} \frac{1}{n}\sum_{k=1}^{n} K(x+y-x_k)m(x_k) - m(x+y)dy + \int_{-1/2n}^{1/2n} m(x+y)dy - m(x) \\
&= b_K^{\pm}(x) + O\left(n^{-3}\right)\max|m''| \qquad (20)
\end{aligned}$$

For the expansion of b_K, we need $\widehat{m}(2\pi\kappa) := \int_0^1 m(x)e^{i2\pi\kappa x}dx$.

$$\begin{aligned}
\widehat{b}_K(2\pi\kappa) &= \int_0^1 \left(\frac{1}{n}\sum_{k=1}^{n} K(x-x_k)m(x_k) - m(x)\right) e^{i2\pi\kappa x}dx \\
&= \int_0^1 \frac{1}{n}\sum_{k=1}^{n} K(x-x_k)m(x_k)e^{i2\pi\kappa x}dx - \widehat{m}(2\pi\kappa)
\end{aligned}$$

$$\begin{aligned}
\widehat{b}_K(2\pi\kappa) + \widehat{m}(2\pi\kappa) &= \int_0^1 \frac{1}{n}\sum_{k=1}^{n} K(x-x_k)m(x_k)e^{i2\pi\kappa x}dx \\
&= \frac{1}{n}\sum_{k=1}^{n}\int_0^1 K(x-x_k)e^{i2\pi\kappa(x-x_k)}dx\ m(x_k)e^{i2\pi\kappa x_k} \\
&= \widehat{K}(2\pi\kappa)\ \frac{1}{n}\sum_{k=1}^{n}\ m(x_k)e^{i2\pi\kappa x_k} \\
&= \widehat{K}(2\pi\kappa)\left(\widehat{m}(2\pi\kappa) + \frac{1}{n}\sum_{k=1}^{n}\ m(x_k)e^{i2\pi\kappa x_k} - \widehat{m}(2\pi\kappa)\right) \\
&= \widehat{\mathscr{b}}_K(2\pi\kappa) + \widehat{m}(2\pi\kappa) + \widehat{K}(2\pi\kappa)\left(\frac{1}{n}\sum_{k=1}^{n}\ m(x_k)e^{i2\pi\kappa x_k} - \widehat{m}(2\pi\kappa)\right)
\end{aligned}$$

Bounding $|(e^{i2\pi\kappa x})'| \leq 2\pi\kappa$ and $|(e^{i2\pi\kappa x})''| \leq (2\pi\kappa)^2$, respectively, $\widehat{K}(2\pi\kappa) \leq \|K\|_2 |2\kappa+1|^{-1/2}$ and $|\kappa| \leq \sqrt{n}$ we obtain:

$$|\widehat{b}_K(2\pi\kappa) - \widehat{\mathscr{b}}_K(2\pi\kappa)| = \widehat{K}(2\pi\kappa)\Big|\frac{1}{n}\sum_{k=1}^{n}\ m(x_k)e^{i2\pi\kappa x_k} - \widehat{m}(2\pi\kappa)\Big|$$

$$
\begin{aligned}
&= \widehat{K}(2\pi\kappa)\frac{1}{n^2}\max_x |(m(x)e^{i2\pi\kappa x})''| \\
&l \quad \widehat{K}(2\pi\kappa)\frac{1}{n^2}\ 4\ (2\pi\kappa)^2 \max\{|m|,|m'|,|m''|\} \\
&= \kappa^2\ \widehat{K}(2\pi\kappa) O\left(n^{-2}\right) \qquad (21) \\
&= O\left(n^{-2}\right)\ \kappa^2\ \frac{\sqrt{\int_0^1 K^2(x)dx}}{\sqrt{2|\kappa|+1}} \\
&= O\left(n^{-5/4}\right)\sqrt{\int_0^1 K^2(x)dx} \qquad (22)
\end{aligned}
$$

Next we aim considering the vector $b_{\mathbb{K}}^T\mathbb{K}$.

$$
\frac{1}{n}\left(b_{\mathbb{K}}^T\mathbb{K}\right)_i = \frac{1}{n}\sum_{k=1}^n b_{\mathbb{K}}(x_k)K^{\pm}(x_i - x_k) = (b_{\mathbb{K}} \diamond K^{\pm})(x_i) \qquad (23)
$$

$$
\begin{aligned}
\left(b_{\mathbb{K}} \diamond K^{\pm}\right)(x) &= \frac{1}{n}\sum_{k=1}^n b_{\mathbb{K}}(x_k)K^{\pm}(x - x_k) \\
&= \sum_{k=1}^n \left[b_K^{\pm}(x_k) + O\left(n^{-3}\right)\right]K^{\pm}(x - x_k) \\
&= n\left(b_K^{\pm} \diamond K\right)^{\pm}(x) + O\left(n^{-3}\right)\sqrt{\int_0^1 K(x)dx}
\end{aligned}
$$

$$
\begin{aligned}
n\widehat{\left(b_K^{\pm} \diamond K\right)}(2\pi\kappa) &= \int_0^1 \sum_{k=1}^n b_K^{\pm}(x_k)K(x - x_k)e^{i2\pi\kappa x}dx \\
&= \sum_{k=1}^n b_K^{\pm}(x_k)\int_0^1 K(x - x_k)e^{i2\pi\kappa(x - x_k)}dx\ e^{i2\pi\kappa x_k} \\
&= \widehat{K}(2\pi\kappa)\sum_{k=1}^n b_K^{\pm}(x_k)\ e^{i2\pi\kappa x_k} \\
&= \widehat{K}(2\pi\kappa)\left(\widehat{b}_K(2\pi\kappa) + \sum_{k=1}^n b_K^{\pm}(x_k)\ e^{i2\pi\kappa x_k} - \widehat{b}_K(2\pi\kappa)\right)
\end{aligned}
$$

Since $y \in [-1/(2n), 1/(2n)]$ we can bound $|e^{i2\pi\kappa x_k} - e^{i2\pi\kappa(x_k+y)}| \leq 2\pi|\kappa|/(2n)$.

$$
\begin{aligned}
\left|n\widehat{\left(b_K^{\pm} \diamond K\right)}(2\pi\kappa) - \widehat{K}(2\pi\kappa)\widehat{b}_K(2\pi\kappa)\right| &= \widehat{K}(2\pi\kappa)\left|\sum_{k=1}^n \int_{-1/2n}^{1/2n} b_K(x_k + y)\left(e^{i2\pi\kappa x_k} - e^{i2\pi\kappa(x_k+y)}\right)dy\right| \\
&\leq \widehat{K}(2\pi\kappa)\sum_{k=1}^n \int_{-1/2n}^{1/2n} |b_K(x_k + y)|dy\ \frac{2\pi|\kappa|}{2n} \\
&\leq \frac{\pi|\kappa|}{n}\ \widehat{K}(2\pi\kappa)\sqrt{\int_0^1 b_K^2(y)dy}
\end{aligned}
$$

$$
\begin{aligned}
&\leq \frac{\pi|\kappa|}{n\sqrt{2|\kappa|+1}} \sqrt{\int_0^1 K^2(x)dx}\sqrt{\int_0^1 b_K^2(y)dy} \\
&= O\left(n^{-3/4}\right) \sqrt{\int_0^1 K^2(x)dx}\sqrt{\int_0^1 b_K^2(y)dy} \qquad (24)
\end{aligned}
$$

Because of $\max|K| \leq 2\sqrt{n}+1$, $\max|K'| \leq 4\pi n$ and $\max|K''| \leq 8\pi^2 n^{3/2}$, (24) is related to MISE in the following way:

$$
\begin{aligned}
|b_K(x) - \mathscr{b}_K(x)| &= \Big|\frac{1}{n}\sum_{k=1}^{n} K(x-x_k)m(x_k) - \int_0^1 K(x-y)m(y)dy\Big| \\
&\leq \frac{1}{n^2}\max_y \left|(K(x-y)m(y))''\right| \\
&= O\left(n^{-1/2}\right)\max\{|m|, |m'|, |m''|\}
\end{aligned}
$$

$$
\begin{aligned}
\Big|\int_0^1 b_K^2(x)dx - \int_0^1 \mathscr{b}_K^2(x)dx\Big| &= \int_0^1 \Big(b_K(x) - \mathscr{b}_K(x)\Big)\Big(2\mathscr{b}_K(x) + b_K(x) - \mathscr{b}_K(x)\Big)dx \\
&= \int_0^1 O\left(n^{-1/2}\right)\left(2\mathscr{b}_K(x) + O\left(n^{-1/2}\right)\right)dx \\
&= O\left(n^{-1/2}\right)\sqrt{\int_0^1 \mathscr{b}_K^2(x)dx} + O\left(n^{-1}\right) \\
&= O(n^{-\delta})MISE(K) + O\Big(n^{\delta-1}\Big)
\end{aligned}
$$

Combining inequalities number (21) and (24) we get:

$$
\begin{aligned}
&\Big|n\big(\widehat{b_K^{\pm} \diamond K}\big)(2\pi\kappa) - \widehat{K}(2\pi\kappa)\widehat{\mathscr{b}}_K(2\pi\kappa)\Big| \\
\leq\; &\Big|n\big(\widehat{b_K^{\pm} \diamond K}\big)(2\pi\kappa) - \widehat{K}(2\pi\kappa)\widehat{b}_K(2\pi\kappa)\Big| + \Big|\widehat{K}(2\pi\kappa)\widehat{b}_K(2\pi\kappa) - \widehat{K}(2\pi\kappa)\widehat{\mathscr{b}}_K(2\pi\kappa)\Big| \\
=\; &O\left(n^{-3/4}\right)\sqrt{\int_0^1 K^2(x)dx}\sqrt{\int_0^1 b_K^2(y)dy} + \widehat{K}^2(2\pi\kappa)O\left(n^{-2}\right)\kappa^2 \\
=\; &O\left(n^{-3/4}\right)\sqrt{\int_0^1 K^2(x)dx}\sqrt{\int_0^1 b_K^2(y)dy} + O\left(n^{-2}\right)\int_0^1 K^2(x)dx\,\frac{\kappa^2}{2|\kappa|+1} \\
=\; &O\left(n^{-3/4}\right)\sqrt{\int_0^1 K^2(x)dx}\sqrt{\int_0^1 b_K^2(y)dy} + O\left(n^{-3/2}\right)\int_0^1 K^2(x)dx \qquad (25)
\end{aligned}
$$

And finally we break down the remaining term $b_{\mathbb{K}}^T\mathbb{C}$ to $(b_{\mathbb{K}}^T\mathbb{C})_i = \frac{1}{2}b_{\mathbb{K}}(x_{i-1}) - b_{\mathbb{K}}(x_i) + \frac{1}{2}b_{\mathbb{K}}(x_{i+1})$.

2.2 The quadratic forms

As in Part I, the quadratic forms (and later on the linear ones) will be decomposed by means of a wavelet basis. The only difference will lie in the replacement of the density function f, by the regression function m, and the fact that sums instead of integrals will appear. Beyond that, the proofs follow the same ideas.

In (12) we have defined

$$\widehat{K}(2\pi\kappa) \;=\; \int_0^1 K(x)e^{i2\pi\kappa x}dx$$

and we recall the assumptions $\widehat{K}(0)=1$, $\widehat{K}(2\pi\kappa)$ monotonously decreasing for $\kappa=1,\dots,\lfloor\sqrt{n}\rfloor$ (where $\lfloor x\rfloor$ is defined to be the integer part of the real number x, and for positive$x$ $\lceil x\rceil = \lfloor x\rfloor+1$) and equal to 0 for $\kappa>\sqrt{n}$. The Fourier coefficients are now to be decomposed by a wavelet basis. Since in Fourier expansion we only need to represent Fourier coefficients defined on $2\pi\mathbb{Z}$, it suffices to employ basis functions which map $2\pi\mathbb{Z}\longrightarrow\mathbb{R}$. Let $\delta_u(.)$ be the Landau symbol, such that $\delta_u(X)=1$, if $X=u$ and 0 otherwise.

$$\begin{aligned}
\widehat{\varphi}_{01}(2\pi\kappa) &:= \delta_0(2\pi\kappa)\\
\widehat{\varphi}_{s2}(2\pi\kappa) &:= 2^{-(s+1)/2}\sum_{u=2^s}^{2^{s+1}-1}\delta_{2\pi u}(|2\pi\kappa|), \qquad s=0,\dots,d_n=\lceil \ln n/(2\ln 2)\rceil\\
\widehat{\psi}_{st}(2\pi\kappa) &:= 2^{-(s+1)/2}\left[\sum_{u=(t-1)2^s}^{(t-1/2)2^s-1}\delta_{2\pi u}(|2\pi\kappa|)-\sum_{u=(t-1/2)2^s}^{t2^s-1}\delta_{2\pi u}(|2\pi\kappa|)\right],\\
&\qquad \text{for } s=1,\dots,d_n;\; t=2,\dots,\lceil 2^{-s}\sqrt{n}\rceil
\end{aligned} \tag{26}$$

We will not need any finer mother wavelets because it is not necessary to represent the functions $\widehat{K}$ in more than $\lfloor\sqrt{n}\rfloor$ arguments. $\{\widehat{\varphi}_{st},\widehat{\psi}_{st}\}$ are orthonormal. They constitute a complete wavelet basis for the functions $\widehat{K}: 2\pi\cdot\{-\lfloor\sqrt{n}\rfloor,-\lfloor\sqrt{n}\rfloor-1,\dots,\lfloor\sqrt{n}\rfloor\}\longrightarrow[0,1]$. The wavelet coefficients are:

$$\alpha_{st}(K) := \sum_{|\kappa|\le\sqrt{n}}\widehat{K}(\kappa)\widehat{\varphi}_{st}(\kappa) \qquad \text{and} \qquad \beta_{st}(K) := \sum_{|\kappa|\le\sqrt{n}}\widehat{K}(\kappa)\widehat{\psi}_{st}(\kappa)$$

The wavelet basis of the Fourier coefficients may be transformed into a function basis of the periodic functions K with period length 1 defined in the time domain,

$$\begin{aligned}
\varphi_{01}(x) &\equiv 1\\
\varphi_{s2}(x) &= 2^{-(s+1)/2}\sum_{|u|=2^s}^{2^{s+1}-1}e^{-i2\pi ux}\\
\psi_{st}(x) &= 2^{-(s+1)/2}\left[\sum_{|u|=(t-1)2^s}^{(t-1/2)2^s-1}e^{-i2\pi ux}-\sum_{|u|=(t-1/2)2^s}^{t2^s-1}e^{-i2\pi ux}\right]
\end{aligned} \tag{27}$$

so that there results a decomposition of K. The sums $\sum_{t=2}^{2^{-s}\sqrt{n}}$ are to be understood as $\sum_{t=2}^{\lceil 2^{-s}\sqrt{n}\rceil}$.

$$\begin{aligned}
K(x) &= \sum_{\kappa\in\mathbb{Z}}\widehat{K}(\kappa)e^{-i2\pi\kappa x}\\
&= \sum_{\kappa\in\mathbb{Z}}\left[\alpha_{01}(K)\widehat{\varphi}_{01}(2\pi\kappa)+\sum_{s=0}^{d_n}\alpha_{s2}(K)\widehat{\varphi}_{s2}(2\pi\kappa)+\sum_{s=1}^{d_n}\sum_{t=2}^{2^{-s}\sqrt{n}}\beta_{st}(K)\widehat{\psi}_{st}(2\pi\kappa)\right]e^{-i2\pi\kappa x}\\
&= \alpha_{01}(K)\sum_{\kappa\in\mathbb{Z}}\widehat{\varphi}_{01}(2\pi\kappa)e^{-i2\pi\kappa x}+\sum_{s=0}^{d_n}\alpha_{s2}(K)\sum_{\kappa\in\mathbb{Z}}\widehat{\varphi}_{s2}(2\pi\kappa)e^{-i2\pi\kappa x}
\end{aligned}$$

$$+\sum_{s=1}^{d_n}\sum_{t=2}^{2^{-s}\sqrt{n}}\beta_{st}(K)\sum_{\kappa\in\mathbb{Z}}\widehat{\psi}_{st}(2\pi\kappa)e^{-i2\pi\kappa x}$$

$$= \alpha_{01}(K)\varphi_{01}(x)+\sum_{s=0}^{d_n}\alpha_{s2}(K)\varphi_{s2}(x)+\sum_{s=1}^{d_n}\sum_{t=2}^{2^{-s}\sqrt{n}}\beta_{st}(K)\psi_{st}(x)$$

Analogously to $\mathbb{K}$, the following matrices are constructed

$$\Phi_{st} := \left(\left(\int_{-1/2n}^{1/2n}\varphi_{st}(x_j-x_i+z)dz\right)\right)_{i,j=1}^{n} \quad \Psi_{st} := \left(\left(\int_{-1/2n}^{1/2n}\psi_{st}(x_j-x_i+z)dz\right)\right)_{i,j=1}^{n} \tag{28}$$

and they thus decompose the quadratic form into:

$$\begin{aligned}\frac{1}{n}\varepsilon^T\mathbb{K}\varepsilon-\frac{1}{n}E\left[\varepsilon^T\mathbb{K}\varepsilon\right] &= \frac{1}{n}\alpha_{01}(K)\left(\varepsilon^T\Phi_{01}\varepsilon-E\left[\varepsilon^T\Phi_{01}\varepsilon\right]\right)+\frac{1}{n}\sum_{s=0}^{d_n}\alpha_{s2}(K)\left(\varepsilon^T\Phi_{s2}\varepsilon-E\left[\varepsilon^T\Phi_{s2}\varepsilon\right]\right)\\ &\quad+\frac{1}{n}\sum_{s=1}^{d_n}\sum_{t=2}^{2^{-s}\sqrt{n}}\beta_{st}(K)\left(\varepsilon^T\Psi_{st}\varepsilon-E\left[\varepsilon^T\Psi_{st}\varepsilon\right]\right)\end{aligned} \tag{29}$$

On a set of “favorable events” $B_{n1,1}$:

$$B_{n1,1} := \left\{(\varepsilon_1,\ldots,\varepsilon_n) : |\varepsilon^T\Phi_{st}\varepsilon-E\left[\varepsilon^T\Phi_{st}\varepsilon\right]|\le\lambda\ln n,\ |\varepsilon^T\Psi_{st}\varepsilon-E\left[\varepsilon^T\Psi_{st}\varepsilon\right]|\le\lambda\ln n,\ \forall\ s,t\right\} \tag{30}$$

with $\lambda<\infty$ sufficiently large, it holds that

$$\begin{aligned}\left|\frac{1}{n}\varepsilon^T\mathbb{K}\varepsilon-\frac{1}{n}E\left[\varepsilon^T\mathbb{K}\varepsilon\right]\right| &\le \frac{1}{n}\Big|\alpha_{01}(K)\left(\varepsilon^T\Phi_{01}\varepsilon-E\left[\varepsilon^T\Phi_{01}\varepsilon\right]\right)+\sum_{s=0}^{d_n}\alpha_{s2}(K)\left(\varepsilon^T\Phi_{s2}\varepsilon-E\left[\varepsilon^T\Phi_{s2}\varepsilon\right]\right)\\ &\quad+\sum_{s=1}^{d_n}\sum_{t=2}^{2^{-s}\sqrt{n}}\beta_{st}(K)\left(\varepsilon^T\Psi_{st}\varepsilon-E\left[\varepsilon^T\Psi_{st}\varepsilon\right]\right)\Big|\\ &\le \frac{\lambda\ln n}{n}\left[|\alpha_{01}(K)|+\sum_{s=0}^{d_n}|\alpha_{st}(K)|+\sum_{s=1}^{d_n}\sum_{t=2}^{2^{-s}\sqrt{n}}|\beta_{st}(K)|\right]\\ &\le \frac{\lambda\ln n}{n}\left[\sqrt{\int_0^1K^2(x)dx}\ \sqrt{d_n+2}+\sum_{s=1}^{d_n}\sqrt{\int_0^1K^2(x)dx}\right]\\ &\le \frac{\lambda\ln n}{n}\sqrt{\int_0^1K^2(x)dx}\ \left(\sqrt{d_n+2}+\sqrt{2}\,d_n\right)\\ &= O\left(\frac{\ln^2 n}{n}\right)\sqrt{\int_0^1K^2(x)dx} \qquad \text{Lemma II.2.1}\end{aligned}$$

This is asymptotically dominated by $O(n^{-\delta})MISE(K)+O(n^{\delta-1})$. In Lemma II.2.2, the probability of a single complementary set, i.e. $P(|\varepsilon^T\Phi_{st}\varepsilon-E\left[\varepsilon^T\Phi_{st}\varepsilon\right]|>\lambda\ln n)$ and $P(|\varepsilon^T\Psi_{st}\varepsilon-E\left[\varepsilon^T\Psi_{st}\varepsilon\right]|>\lambda\ln n)$, respectively, is calculated, and for the union we obtain

$$P\left(B_{n1,1}^c\right) \le P\left(|\varepsilon^T\Phi_{01}\varepsilon-E\left[\varepsilon^T\Phi_{01}\varepsilon\right]|>\lambda\ln n\right)+\sum_{s=0}^{d_n}P\left(|\varepsilon^T\Phi_{s2}\varepsilon-E\left[\varepsilon^T\Phi_{s2}\varepsilon\right]|>\lambda\ln n\right)$$

$$
\begin{aligned}
& + \sum_{s=1}^{d_n} \sum_{t=2}^{2^{-s}\sqrt{n}} P\left(\left|\varepsilon^T \Psi_{st} \varepsilon - E\left[\varepsilon^T \Psi_{st} \varepsilon\right]\right| > \lambda \ln n\right) \\
= \; & (d_n + 2) O\left(n^{-\lambda}\right) + \sum_{s=1}^{d_n} \left(\lceil 2^{-s}\sqrt{n} \rceil - 1\right) O\left(n^{-\lambda}\right) \\
= \; & O\left(n^{-\lambda + 1/2}\right)
\end{aligned}
$$

Applying (16) we decompose $\varepsilon^T \mathbb{K}^2 \varepsilon - E[\varepsilon^T \mathbb{K}^2 \varepsilon]$ trough the same $\{\varphi_{st}, \psi_{st}\}$. And on the set $B_{n1,1}$ we obtain:

$$
\begin{aligned}
& \left| \frac{1}{n} \varepsilon^T \mathbb{K}^2 \varepsilon - \frac{1}{n} E\left[\varepsilon^T \mathbb{K}^2 \varepsilon\right] \right| \\
= \; & \left| \frac{1}{n} \varepsilon^T n \left(K \diamond K^{\pm}\right) \varepsilon - E\left[\frac{1}{n} \varepsilon^T n \left(K \diamond K^{\pm}\right) \varepsilon\right] \right| \\
= \; & \frac{1}{n} \Big| \alpha_{01}(n\left(K \diamond K^{\pm}\right)) \left(\varepsilon^T \Phi_{01} \varepsilon - E\left[\varepsilon^T \Phi_{01} \varepsilon\right]\right) + \sum_{s=0}^{d_n} \alpha_{s2}(n\left(K \diamond K^{\pm}\right)) \left(\varepsilon^T \Phi_{s2} \varepsilon - E\left[\varepsilon^T \Phi_{s2} \varepsilon\right]\right) \\
& + \sum_{s=1}^{d_n} \sum_{t=2}^{2^{-s}\sqrt{n}} \beta_{st}(n\left(K \diamond K^{\pm}\right)) \left(\varepsilon^T \Psi_{st} \varepsilon - E\left[\varepsilon^T \Psi_{st} \varepsilon\right]\right) \Big| \\
= \; & \frac{\lambda \ln n}{n} \left[\sqrt{\int_0^1 K^2(x) dx} \sqrt{d_n + 2} + O\left(n^{-1/2}\right) \int_0^1 K^2(x) dx \right. \\
& \left. + \sum_{s=1}^{d_n} \sqrt{\int_0^1 K^2(x) dx} + 2^{-s/2}\, O\left(n^{-1/4}\right) \int_0^1 K^2(x) dx \right] \\
= \; & O\left(\frac{\ln^2 n}{n}\right) \sqrt{\int_0^1 K^2(x) dx} + O\left(n^{-5/4} \ln n\right) \int_0^1 K^2(x) dx \qquad \text{Lemma II.2.1}
\end{aligned}
$$

Next we consider the term $\varepsilon^T \mathbb{K}\mathbb{C} \varepsilon - E[\varepsilon^T \mathbb{K}\mathbb{C} \varepsilon]$. Here it is useful to construct new basis matrices out of the old basis functions.

$$
\begin{aligned}
\Theta_{st} \; &:= \; \left(\left(\frac{1}{2} \int_{1/2n}^{3/2n} \varphi_{st}(x_i - x_j + z) + \varphi_{st}(x_i - x_j - z) dz - \int_{-1/2n}^{1/2n} \varphi_{st}(x_i - x_j + z) dz \right) \right)_{i,j=1}^{n} \\
\Xi_{st} \; &:= \; \left(\left(\frac{1}{2} \int_{1/2n}^{3/2n} \psi_{st}(x_i - x_j + z) + \psi_{st}(x_i - x_j - z) dz - \int_{-1/2n}^{1/2n} \psi_{st}(x_i - x_j + z) dz \right) \right)_{i,j=1}^{n} \qquad (31)
\end{aligned}
$$

The wavelet coefficients correspond to those of K, and Lemma II.2.1 may be applied again. The quadratic forms however, are not the same as for $\mathbb{K}$. Hence it is necessary to construct a new set of "favorable events" $B_{n1,2}$, where the quadratic forms in Θ_{st} and Ξ_{st} do not exceed $\lambda \ln n$. The bound for $P(B^c_{n1,2})$ can be found analogously to the one for $P(B^c_{n1,1})$. On the set $B_{n1,2}$ we have

$$
\left| \frac{1}{n} \varepsilon^T \mathbb{K}\mathbb{C} \varepsilon - \frac{1}{n} E\left[\varepsilon^T \mathbb{K}\mathbb{C} \varepsilon\right] \right| \; = \; \frac{1}{n} \Big| \alpha_{01}(K) \left(\varepsilon^T \Theta_{01} \varepsilon - E\left[\varepsilon^T \Theta_{01} \varepsilon\right]\right) + \sum_{s=0}^{d_n} \alpha_{s2}(K) \left(\varepsilon^T \Theta_{s2} \varepsilon - E\left[\varepsilon^T \Theta_{s2} \varepsilon\right]\right)
$$

$$
\begin{aligned}
& + \sum_{s=1}^{d_n} \sum_{t=2}^{2^{-s}\sqrt{n}} \beta_{st}(K) \left(\varepsilon^T \Xi_{st} \varepsilon - E\left[\varepsilon^T \Xi_{st} \varepsilon \right] \right) \Big| \\
\leq \ & \frac{\lambda \ln n}{n} \left(|\alpha_{01}(K)| + \sum_{s=0}^{d_n} |\alpha_{s2}(K)| + \sum_{s=1}^{d_n} \sum_{t=2}^{2^{-s}\sqrt{n}} |\beta_{st}(K)| \right) \\
= \ & O\left(\frac{\ln^2 n}{n} \right) \sqrt{\int_0^1 K^2(x)dx} \qquad (32)
\end{aligned}
$$

$$
\begin{aligned}
P\left(B^c_{n1,2}\right) \ & \leq \ (d_n+2) O\left(n^{-\lambda}\right) + \sum_{s=1}^{d_n} \left(2^{-s} n - 1\right) O\left(n^{-\lambda}\right) \\
& = \ O\left(n^{-\lambda+1/2}\right) \qquad \text{Lemma II.2.2}
\end{aligned}
$$

There still remain the terms $\varepsilon^T \mathbb{C} \varepsilon - E\left[\varepsilon^T \mathbb{C} \varepsilon\right]$ and $\varepsilon^T \mathbb{C}^2 \varepsilon - E\left[\varepsilon^T \mathbb{C}^2 \varepsilon\right]$ to regard upon. The matrices do not depend on K and therefore no wavelet decomposition is needed. Our aim is to find a bound for both quadratic forms that is only exceeded with a rapidly vanishing probability. According to Lemma II.2.2, this time the bound falls at $\lambda\sqrt{n \ln n}$. The sets of "favorable events" can be denoted by $B_{n2,1}$ for $\mathbb{C}$ and $B_{n2,2}$ for $\mathbb{C}^2$. On these sets

$$
\frac{1}{n\sqrt{n}} \left(\varepsilon^T \mathbb{C} \varepsilon - E\left[\varepsilon^T \mathbb{C} \varepsilon\right] \right) \ = \ O\left(\frac{\lambda \ln^{1/2} n}{n} \right) \qquad (33)
$$

$$
P\left(B^c_{n2,1}\right) \ \leq \ O\left(n^{-\lambda}\right)
$$

and

$$
\frac{1}{n^2} \left(\varepsilon^T \mathbb{C}^2 \varepsilon - E\left[\varepsilon^T \mathbb{C}^2 \varepsilon\right] \right) \ = \ O\left(\frac{\lambda \ln^{1/2} n}{n\sqrt{n}} \right) \qquad (34)
$$

$$
P\left(B^c_{n2,2}\right) \ \leq \ O\left(n^{-\lambda}\right)
$$

so that these terms are dominated by $O(n^{-\delta}) MISE(K) + O(n^{\delta-1})$, too, for arbitrary δ.

Lemma II.2.1

$$
\begin{aligned}
|\alpha_{01}(K)| + \sum_{s=0}^{d_n} |\alpha_{s2}(K)| \ & \leq \ \sqrt{\int_0^1 K^2(x)dx} \ \sqrt{d_n+2} \\
\sum_{t=2}^{2^{-s}\sqrt{n}} |\beta_{st}(K)| \ & \leq \ \sqrt{\int_0^1 K^2(x)dx}
\end{aligned}
$$

and analogously

$$
\begin{aligned}
|\alpha_{01}(n\left(K \diamond K^{\pm}\right))| + \sum_{s=0}^{d_n} |\alpha_{s2}(n\left(K \diamond K^{\pm}\right))| \ & \leq \ \sqrt{\int_0^1 K^2(x)dx} \ \sqrt{d_n+2} + O\left(n^{-1/2}\right) \int_0^1 K^2(x)dx \\
\sum_{t=2}^{2^{-s}\sqrt{n}} |\beta_{st}(n\left(K \diamond K^{\pm}\right))| \ & \leq \ \sqrt{\int_0^1 K^2(x)dx + 2^{-s/2} O\left(n^{-1/4}\right) \int_0^1 K^2(x)dx}
\end{aligned}
$$

Proof Because of orthonormality we can write:

$$\begin{aligned}
\left[|\alpha_{01}(K)|+\sum_{s=0}^{d_n}|\alpha_{s2}(K)|\right] &= \Big|\sum_{|\kappa|\le\sqrt{n}}\widehat{K}(2\pi\kappa)\widehat{\varphi}_{01}(2\pi\kappa)\Big|+\sum_{s=0}^{d_n}\Big|\sum_{|\kappa|\le\sqrt{n}}\widehat{K}(2\pi\kappa)\widehat{\varphi}_{s2}(2\pi\kappa)\Big| \\
&= \sum_{|\kappa|\le\sqrt{n}}\widehat{K}(2\pi\kappa)\left[\widehat{\varphi}_{01}(2\pi\kappa)+\sum_{s=0}^{d_n}\widehat{\varphi}_{s2}(2\pi\kappa)\right] \\
&\le \sqrt{\sum_{|\kappa|\le\sqrt{n}}\widehat{K}^2(2\pi\kappa)}\sqrt{\sum_{|\kappa|\le\sqrt{n}}\left(\widehat{\varphi}_{01}(2\pi\kappa)+\sum_{s=0}^{d_n}\widehat{\varphi}_{s2}(2\pi\kappa)\right)^2} \\
&= \sqrt{\int_0^1 K^2(x)dx}\sqrt{\sum_{|\kappa|\le\sqrt{n}}\left[\widehat{\varphi}^2_{01}(2\pi\kappa)+\sum_{s=0}^{d_n}\widehat{\varphi}^2_{s2}(2\pi\kappa)\right]} \\
&= \sqrt{\int_0^1 K^2(x)dx}\sqrt{1+\sum_{s=0}^{d_n}1} \\
&= \sqrt{\int_0^1 K^2(x)dx}\ \sqrt{d_n+2}
\end{aligned}$$

For the β's we use the known bound $\widehat{K}(2\pi\kappa)\le\|K\|_2|2\kappa+1|^{-1/2}$.

$$\begin{aligned}
\sum_{t=2}^{2^{-s}\sqrt{n}}|\beta_{st}(K)| &= \sum_{t=2}^{2^{-s}\sqrt{n}}\Big|\sum_{|\kappa|\le\sqrt{n}}\widehat{K}(2\pi\kappa)\widehat{\psi}_{st}(2\pi\kappa)\Big| \\
&= \sum_{t=2}^{2^{-s}\sqrt{n}}\Big|\sum_{|\kappa|\le\sqrt{n}}\widehat{K}(2\pi\kappa)2^{-(s+1)/2}\left[\sum_{u=(t-1)2^s}^{(t-1/2)2^s-1}\delta_u(|2\pi\kappa|)-\sum_{u=(t-1/2)2^s}^{t2^s-1}\delta_u(|2\pi\kappa|)\right]\Big| \\
&= 2^{-(s+1)/2}\sum_{t=2}^{2^{-s}\sqrt{n}}\Big|2\sum_{u=(t-1)2^s}^{(t-1/2)2^s-1}\widehat{K}(2\pi u)-2\sum_{u=(t-1/2)2^s}^{t2^s-1}\widehat{K}(2\pi u)\Big| \\
&\le 2^{-(s-1)/2}\sum_{t=2}^{2^{-s}\sqrt{n}}\left(2^s\widehat{K}\left(2\pi(t-1)2^s\right)-2^s\widehat{K}\left(2\pi t2^s\right)\right) \\
&\le 2^{(s+1)/2}\widehat{K}\left(2\pi 2^s\right) \\
&\le 2^{(s+1)/2}\frac{\sqrt{\sum_{|\kappa|\le\sqrt{n}}\widehat{K}^2(2\pi\kappa)}}{\sqrt{2\cdot 2^s+1}} \\
&\le \sqrt{\int_0^1 K^2(x)dx}
\end{aligned}$$

To obtain the bounds for the coefficients of $n(K\diamond K^{\pm})$ we revert to equation (17).

$$|\alpha_{01}(n\left(K\diamond K^{\pm}\right))|+\sum_{s=0}^{d_n}|\alpha_{s2}(n\left(K\diamond K^{\pm}\right))|$$

$$
\begin{aligned}
&\le \sum_{|\kappa|\le\sqrt{n}} \left| n\,(\widehat{K\diamond K^{\pm}})(2\pi\kappa)\right| \left[\widehat{\varphi}_{01}(2\pi\kappa)+\sum_{s=0}^{d_n}\widehat{\varphi}_{s2}(2\pi\kappa)\right] \\
&= \sum_{|\kappa|\le\sqrt{n}} \left(\widehat{K}^2(2\pi\kappa)+O\left(n^{-3/4}\right)\int_0^1 K^2(x)dx\right)\left[\widehat{\varphi}_{01}(2\pi\kappa)+\sum_{s=0}^{d_n}\widehat{\varphi}_{s2}(2\pi\kappa)\right],
\end{aligned}
$$

where the term $\sum \widehat{K}^2(2\pi\kappa)[\widehat{\varphi}_{01}(2\pi\kappa)+\sum_{s=0}^{d_n}\widehat{\varphi}_{s2}(2\pi\kappa)]$ is handled like $\sum \widehat{K}(2\pi\kappa)[\widehat{\varphi}_{01}(2\pi\kappa)+\sum_{s=0}^{d_n}\widehat{\varphi}_{s2}(2\pi\kappa)]$. The other one is bounded by

$$
\begin{aligned}
&\sum_{|\kappa|\le\sqrt{n}} O\left(n^{-3/4}\right)\int_0^1 K^2(x)dx\left[\widehat{\varphi}_{01}(2\pi\kappa)+\sum_{s=0}^{d_n}\widehat{\varphi}_{s2}(2\pi\kappa)\right] \\
&= O\left(n^{-3/4}\right)\int_0^1 K^2(x)dx\left[1+\sum_{s=0}^{d_n}2^{-(s+1)/2}\sum_{|\kappa|=2^s}^{2^{s+1}-1}1\right] \\
&= O\left(n^{-3/4}\right)\int_0^1 K^2(x)dx\left[1+\sum_{s=0}^{d_n}2^{(s+1)/2}\right] \\
&= O\left(n^{-1/2}\right)\int_0^1 K^2(x)dx,
\end{aligned}
$$

since $2^{d_n/2}=O(n^{1/4})$. A similar calculation leads to the desired result for $\sum|\beta_{st}(n(K\diamond K^{\pm}))|$.

$$
\sum_{t=2}^{2^{-s}\sqrt{n}} |\beta_{s2}(n\left(K\diamond K^{\pm}\right))| \;\le\; \sum_{|\kappa|\le\sqrt{n}}\left(\widehat{K}^2(2\pi\kappa)+O\left(n^{-3/4}\right)\int_0^1 K^2(x)dx\right)\sum_{t=2}^{2^{-s}\sqrt{n}}|\widehat{\psi}_{st}(2\pi\kappa)|
$$

$$
\begin{aligned}
\sum_{|\kappa|\le\sqrt{n}} O\left(n^{-3/4}\right)\int_0^1 K^2(x)dx\sum_{t=2}^{2^{-s}\sqrt{n}}|\widehat{\psi}_{st}(2\pi\kappa)| &= O\left(n^{-3/4}\right)\int_0^1 K^2(x)dx\;2^{-(s+1)/2}\,2\sum_{\kappa=2^s}^{\sqrt{n}}1 \\
&= 2^{-s/2}\,O\left(n^{-1/4}\right)\int_0^1 K^2(x)dx \qquad \square
\end{aligned}
$$

Lemma II.2.2

$$
\begin{aligned}
P\left(|\varepsilon^T\Phi_{st}\varepsilon-E\varepsilon^T\Phi_{st}\varepsilon|>\lambda\ln n\right) &= O\left(n^{-\lambda}\right),\ \text{for } s=0,t=1 \text{ and } s=0,\dots,d_n,t=2 \\
P\left(|\varepsilon^T\Psi_{st}\varepsilon-E\varepsilon^T\Psi_{st}\varepsilon|>\lambda\ln n\right) &= O\left(n^{-\lambda}\right),\ \text{for } s=1,\dots d_n \text{ and } t=2,\dots,2^{-s}\sqrt{n}
\end{aligned}
$$

The same holds for the Θ's and the Ξ's. Furthermore:

$$
\begin{aligned}
P\left(|\varepsilon^T\mathbb{C}\varepsilon-E\varepsilon^T\mathbb{C}\varepsilon|>\lambda\sqrt{n\ln n}\right) &= O\left(n^{-\lambda}\right) \\
P\left(|\varepsilon^T\mathbb{C}^2\varepsilon-E\varepsilon^T\mathbb{C}^2\varepsilon|>\lambda\sqrt{n\ln n}\right) &= O\left(n^{-\lambda}\right)
\end{aligned}
$$

Proof For a matrix A and independent, standard-normally distributed errors ε_i, Lemma 2.1 of Bentkus, Rudzkis (1980) (see also Appendix A.3), holds:

$$
P\left(|\varepsilon^TA\varepsilon-E\varepsilon^TA\varepsilon|>x\right) \;\le\; 2\exp\left\{-\frac{x^2/2}{2\mathrm{tr}\,A^2+2x\|A\|_\infty}\right\}
$$

We apply this inequality to the matrix Φ_{st}.

$$\begin{aligned}
\operatorname{tr}\Phi_{st}^2 &= \sum_{i,j}\left(\int_{-1/2n}^{1/2n}\varphi_{st}(x_j-x_i+y)dy\right)^2 \\
&\leq \sum_{i,j}\frac{1}{n}\int_{-1/2n}^{1/2n}\varphi_{st}^2(x_j-x_i+y)dy \\
&= \sum_{i=1}^{n}\frac{1}{n}\int_0^1\varphi_{st}^2(y)dy \\
&= 1
\end{aligned}$$

$$\begin{aligned}
\|\Phi_{st}\|_\infty &= \max_i\sum_{j=1}^{n}\left|\int_{-1/2n}^{1/2n}\varphi_{st}(x_j-x_i+y)dy\right| \\
&\leq \max_i\sum_{j=1}^{n}\sqrt{\frac{1}{n}}\sqrt{\int_{-1/2n}^{1/2n}\varphi_{st}^2(x_j-x_i+y)dy} \\
&\leq \max_i\sqrt{\sum_{j=1}^{n}\frac{1}{n}}\sqrt{\sum_{j=1}^{n}\int_{-1/2n}^{1/2n}\varphi_{st}^2(x_j-x_i+y)dy} \\
&= \sqrt{\int_0^1\varphi_{st}^2(y)dy} \\
&= 1
\end{aligned}$$

Hence it holds that

$$\begin{aligned}
P\left(|\varepsilon^T\Phi_{st}\varepsilon-E\varepsilon^T\Phi_{st}\varepsilon|>\lambda\ln n\right) &\leq 2\exp\left\{-\frac{\lambda^2\ln^2 n}{4\operatorname{tr}\Phi_{st}^2+4\lambda\ln n\|\Phi_{st}\|_\infty}\right\} \\
&\leq \exp\left\{-\frac{\lambda^2\ln^2 n}{4+4\lambda\ln n}\right\} \\
&= O\Big(\exp\{-\lambda\ln n\}\Big)
\end{aligned}$$

With $\operatorname{tr}\Theta_{st}^2\leq\frac{9}{2}\operatorname{tr}\Phi_{st}^2$ and $\|\Theta_{st}\|_\infty\leq 2\|\Phi_{st}\|_\infty$ the same rate can be achieved for $P(|\varepsilon^T\Theta\varepsilon-E[\varepsilon^T\Theta\varepsilon]|>\lambda\ln n)$. The computations for Ψ_{st} and Ξ_{st} are analogous.
For $\mathbb{C}$ we have $\operatorname{tr}\mathbb{C}^2=\frac{3}{2}n$, $\|\mathbb{C}\|_\infty=2$ and $\operatorname{tr}\mathbb{C}^4=\frac{35}{8}n$, $\|\mathbb{C}^2\|_\infty=4$. In order to avoid losing the exponential rate of $P(B^c_{n2,1})$ and $P(B^c_{n2,2})$, we can only set a bound of $\lambda\sqrt{n\ln n}$ in the inequality of Bentkus, Rudzkis (1980). □

2.3 The linear forms

The basis functions for the decomposition of the low-frequency component of the bias ℓ_K+h_m are constructed in a similar way as in the density estimation case. The regression function m will replace the density function f and instead of integrals we will consider sums. An extra summand added to $\widehat{m}(2\pi\kappa)$ will help us to bound the number of mother wavelets, that we need for our basis.

$$\widehat{m}_n(0):=\widehat{m}(0)+n^{-5/4}\quad\text{and}\quad\widehat{m}_n(2\pi\kappa):=\begin{cases}\widehat{m}(2\pi\kappa)+n^{-5/4}\kappa^{-1}, & \text{if }\widehat{m}(2\pi\kappa)>0\\ \widehat{m}(2\pi\kappa)-n^{-5/4}\kappa^{-1}, & \text{if }\widehat{m}(2\pi\kappa)<0\\ 0, & \text{if }\widehat{m}(2\pi\kappa)=0\end{cases}$$

$$M_n \;:=\; \sum_{|\kappa|\le\sqrt{n}} |\widehat{m}_n(2\pi\kappa)|^2 \tag{35}$$

Since the division of $2\pi\cdot\{-\lfloor\sqrt{n}\rfloor,-\lfloor\sqrt{n}\rfloor+1,\ldots,\lfloor\sqrt{n}\rfloor\}$ into sections Δ_{st} with exactly $\sum_{\Delta_{st}}|\widehat{m}_n(2\pi\kappa)|^2 = 2^{-s}M_n$ is not possible, the basis functions must be normalized and orthogonalized appropriately. Let us define κ_{st}, in case it exists, as:

$$\kappa_{st} \;:=\; \max\left\{\kappa\in\mathbb{N}\;\middle|\; 2^{-s}(t-1)M_n \le 2\sum_{|\lambda|\le\kappa}|\widehat{m}_n(2\pi\lambda)|^2 < 2^{-s}tM_n\right\} \qquad \text{and}$$

$$\Delta_{st}(2\pi\kappa) \;:=\; \sum_{u=\kappa_{s,t-1}+1}^{\kappa_{st}} \delta_u(|2\pi\kappa|) \tag{36}$$

so that we have the following basis functions:

$$\begin{aligned}
\widehat{\varphi'}_{st}(2\pi\kappa) \;&:=\; 2^{s/2}\nu_{st}^{1/2}M_n^{-1/2}\,\widehat{m}_n(2\pi\kappa)\Delta_{st}(2\pi\kappa), \qquad \text{for } s=1,\ldots,s_n:=\lceil\ln n/\ln 2\rceil,\\
&\qquad t=2^s-1 \text{ and for } s=s_n \text{ also } t=2^{s_n}\\
\widehat{\psi'}_{st}(2\pi\kappa) \;&:=\; 2^{(s+1)/2}\mu_{st}M_n^{-1/2}\,\widehat{m}_n(2\pi\kappa)\Big[\nu_{s+1,2t-1}\Delta_{s+1,2t-1}(2\pi\kappa)-\nu_{s+1,2t}\Delta_{s+1,2t}(2\pi\kappa)\Big],\\
&\qquad \text{for } s=1,\ldots,s_n-1; t=2,\ldots,2^s-1 \text{ and } s\ge s_n; t=1,\ldots,2^s
\end{aligned} \tag{37}$$

with appropriate normalization constants:

$$\nu_{st} := \frac{M_n}{2^s\sum_{\kappa}|\widehat{m}_n(2\pi\kappa)|^2\Delta_{st}(2\pi\kappa)} \qquad \text{and} \qquad \mu_{st} := \frac{1}{\sqrt{\nu_{s+1,2t-1}+\nu_{s+1,2t}}} \tag{38}$$

where $\nu_{st}\in[1,2)$ and, correspondingly, $\mu_{st}\in(\frac{1}{2},\frac{1}{\sqrt{2}}]$. Now a basis may be formed by means of $\{\widehat{\varphi}'_{1,1},\widehat{\varphi}'_{2,3},\ldots,\widehat{\varphi}'_{s_n,2^{s_n}-1},\widehat{\varphi}'_{s_n,2^{s_n}}\}$ and $\{\widehat{\psi}'_{s,t}|s=1,\ldots,s_n-1;t=1,\ldots,2^s-1$ and $s=s_n$, $t=2^{s_n}\}$ (as far as they are defined through κ_{st}) and as many finer mother wavelet functions $\widehat{\psi}'_{s,t}$, $s>s_n$, $1\le t\le 2^s$, as needed for guaranteeing that every function $\widehat{m}_n\cdot\widehat{g}$ in the range of $|\kappa|\le\sqrt{n}$ can be exactly represented. The number of wavelets, which must be constructed over a certain point $2\pi\kappa$ depends on the value of $|\widehat{m}_n(2\pi\kappa)|^2$ compared to M_n. It is bounded by the integer $s_\kappa=\min\{s\in\mathbb{N}|2^{-s}M_n\le 2|\widehat{m}_n(2\pi\kappa)|^2\}\le(\ln M_n+5/2\ln n+2\ln\kappa)/\ln 2$. The functions $\{\widehat{\varphi}'_{st},\widehat{\psi}'_{st}\}$ constitute a complete orthonormal basis for all functions $\widehat{g}\cdot\widehat{m}\cdot I_{\sqrt{n}}$ (the same class as $\widehat{g}\cdot\widehat{m}_n\cdot I_{\sqrt{n}}$, because they share their zeros), to which $\widehat{b}_K\cdot I_{\sqrt{n}}$ belong for every $K\in\mathcal{K}'$. With their help, we may start to consider the linear forms. First we make use of the inequality (18), Section 2.1, and $\frac{1}{n}h_m(x)=h_m^{\pm}(x)+o(n^{-3})\max|m''|$:

$$\begin{aligned}
\frac{1}{n}(b_{\mathbb{K}}+h_m)^T\varepsilon \;&=\; \frac{1}{n}\sum_{i=1}^n\Big(b_{\mathbb{K}}(x_i)+h_m(x_i)\Big)\varepsilon_i\\
&=\; \sum_{i=1}^n(b_K+h_m)^{\pm}(x_i)\;\varepsilon_i+O\left(n^{-3}\right)\sum_{i=1}^n|\varepsilon_i|
\end{aligned} \tag{39}$$

The second term of the right-hand side is independent of K, $\sum|\varepsilon_i|$ will be bounded on a set of “favorable events” $B_{n3,0}$ by $O(n)$. The complement of this set has exponentially decreasing probability, which we will see at the end of the section in equation (48).
For $\sum(b_K+h_m)^{\pm}(x_i)\varepsilon_i$ we proceed similarly to the case of density estimation by decomposing b_K+h_m using the basis $\{\varphi'_{st},\psi'_{st}\}$. On the set of “favorable events”

$$B_{n3,1} := \left\{(\varepsilon_1,\ldots,\varepsilon_n) : |\boldsymbol{\varphi}'^{\,T}_{st}\varepsilon| \le \frac{\lambda\ln^{1/2}n}{\sqrt{n}},\ \forall s,t \;\;\text{and}\;\; |\boldsymbol{\psi}'^{\,T}_{st}\varepsilon| \le \frac{\lambda\ln^{1/2}n}{\sqrt{n}},\ \forall s,t\right\} \tag{40}$$

$$\text{with} \quad \boldsymbol{\varphi}'_{st} := \left(\int_{-1/2n}^{1/2n} \varphi'_{st}(x_i + z) dz \right)_{i=1}^{n} \quad \text{and} \quad \boldsymbol{\psi}'_{st} := \left(\int_{-1/2n}^{1/2n} \psi'_{st}(x_i + z) dz \right)_{i=1}^{n}$$

it holds that (remember $\widehat{b}_K(2\pi\kappa) + \widehat{h}_m(2\pi\kappa) = \widehat{b}_K(2\pi\kappa) I_{\sqrt{n}}(\kappa)$):

$$\begin{aligned}
& \left| (b_K + h_m)^{\pm T} \varepsilon \right| \\
= & \left| \left[\sum_{s=1}^{s_n} \alpha'_{s,2^s-1}(b_K) \boldsymbol{\varphi}'_{s,2^s-1} + \alpha'_{s_n,2^{s_n}}(b_K) \boldsymbol{\varphi}'_{s_n,2^{s_n}} + \sum_{s=1}^{s_n-1} \sum_{t=1}^{2^s-1} \beta'_{st}(b_K) \boldsymbol{\psi}'_{st} \right.\right. \\
& \left.\left. + \sum_{s=s_n}^{\max s_\kappa} \sum_{t=1}^{2^s} \beta'_{st}(b_K) \boldsymbol{\psi}'_{st} \right]^T \varepsilon \right| \\
= & \left| \sum_{s=1}^{s_n} \alpha'_{s,2^s-1}(b_K) \left[\boldsymbol{\varphi}'^T_{s,2^s-1} \varepsilon \right] + \alpha'_{s_n,2^{s_n}}(b_K) \left[\boldsymbol{\varphi}'^T_{s_n,2^{s_n}} \varepsilon \right] + \sum_{s=1}^{s_n-1} \sum_{t=1}^{2^s-1} \beta'_{st}(b_K) \left[\boldsymbol{\psi}'^T_{st} \varepsilon \right] \right. \\
& \left. + \sum_{s=s_n}^{\max s_\kappa} \sum_{t=1}^{2^s} \beta'_{st}(b_K) \left[\boldsymbol{\psi}'^T_{st} \varepsilon \right] \right| \\
\leq & \frac{\lambda \ln^{1/2} n}{\sqrt{n}} \left[\sum_{s=1}^{s_n} |\alpha'_{s,2^s-1}(b_K)| + |\alpha'_{s_n,2^{s_n}}(b_K)| + \sum_{s=1}^{s_n-1} \sum_{t=1}^{2^s-1} |\beta'_{st}(b_K)| + \sum_{s=s_n}^{\max s_\kappa} \sum_{t=1}^{2^s} |\beta'_{st}(b_K)| \right] \\
\leq & \frac{\lambda \ln^{1/2} n}{\sqrt{n}} \left[\sqrt{\int_0^1 b_K^2(x) dx} \sqrt{s_n + 1} + O\left(n^{-3/2} \ln n \right) \sqrt{\int_0^1 K^2(x) dx} \right. \\
& \left. + \sum_{s=1}^{s_n-1} \sqrt{2 \int_0^1 b_K^2(x) dx} + \sum_{s=s_n}^{\max s_\kappa} 2 \cdot 2^{-s/2} \|m\|_2 + O\left(n^{-3/4} \right) \sqrt{\int_0^1 K^2(x) dx} + O\left(n^{-1} \right) \right] \\
= & O\left(\frac{\ln^{3/2} n}{\sqrt{n}} \right) \sqrt{\int_0^1 b_K^2(x) dx} + O\left(n^{-5/4} \ln^{1/2} n \right) \sqrt{\int_0^1 K^2(x) dx} + O\left(n^{-1} \ln^{1/2} n \right) \left(\|m\|_2 + 1 \right) \\
= & O\left(\frac{\ln^{3/2} n}{\sqrt{n}} \right) \sqrt{\int_0^1 b_K^2(x) dx} \qquad \text{Lemma II.2.3}
\end{aligned}$$

Obviously $O(n^{-5/4} \ln^{1/2} n) \|K\|_2 = O(n^{-1/4} \ln^{1/2} n) \frac{1}{n} \|K\|_2$ is of smaller order than $O(n^{-1/4} \cdot \ln^{1/2} n) MISE(K)$. The probability of the complementary set can be shown to have the bound (Lemma II.2.4 and (47)):

$$P\left(B^c_{n3,1} \right) = O\left(n^{-\lambda^2 + 1/2} \right) \tag{41}$$

For $\frac{1}{n} b_{\mathbb{K}}^T \mathbb{K} \varepsilon$, we remember (23) of Subsection 2.1. This time, subtraction of h_m from $b_{\mathbb{K}}$ is not necessary because multiplication by $\mathbb{K}$ already suppresses the high frequency component.

$$\begin{aligned}
\frac{1}{n} b_{\mathbb{K}}^T \mathbb{K} \varepsilon &= \sum_{i=1}^{n} (b_{\mathbb{K}} \diamond K^{\pm})(x_i) \varepsilon_i \\
&= \sum_{i=1}^{n} n (b_K^{\pm} \diamond K)^{\pm}(x_i) \varepsilon_i + O\left(n^{-3} \right) \sqrt{\int_0^1 K^2(x) dx} \sum_{i=1}^{n} |\varepsilon_i| \qquad (42)
\end{aligned}$$

Even with the extra factor $\|K\|_2$ the second term is negligible compared to MISE. $n(b_K^{\pm} \diamond K)^{\pm}$ may now be decomposed analogously to $b_K^{\pm}$ through the basis $\{\varphi'_{st}, \psi'_{st}\}$. The set of "favorable events" is thus again $B_{n3,1}$, and the coefficients are handled in Lemma II.2.3. On the set $B_{n3,1}$ it holds that:

$$\begin{aligned}
&\left| n(b_K^{\pm} \diamond K)^{\pm T} \varepsilon \right| \\
= &\left| \left[\sum_{s=1}^{s_n} \alpha'_{s,2^s-1}(n\,(b_K^{\pm} \diamond K))\boldsymbol{\varphi}'_{s,2^s-1} + \alpha'_{s_n,2^{s_n}}(n\,(b_K^{\pm} \diamond K))\boldsymbol{\varphi}'_{s_n,2^{s_n}} + \sum_{s=1}^{s_n-1}\sum_{t=1}^{2^s-1} \beta'_{st}(n\,(b_K^{\pm} \diamond K))\boldsymbol{\psi}'_{st} \right.\right.\\
&\left.\left. + \sum_{s=s_n}^{\max s_\kappa}\sum_{t=1}^{2^s} \beta'_{st}(n\,(b_K^{\pm} \diamond K))\boldsymbol{\psi}'_{st} \right]^T \varepsilon \right| \\
\le &\frac{\lambda \ln^{1/2} n}{\sqrt{n}} \left[\sum_{s=1}^{s_n} |\alpha'_{s,2^s-1}(n\,(b_K^{\pm} \diamond K))| + |\alpha'_{s_n,2^{s_n}}(n\,(b_K^{\pm} \diamond K))| + \sum_{s=1}^{s_n-1}\sum_{t=1}^{2^s-1} |\beta'_{st}(n\,(b_K^{\pm} \diamond K))| \right.\\
&\left. + \sum_{s=s_n}^{\max s_\kappa}\sum_{t=1}^{2^s} |\beta'_{st}(n\,(b_K^{\pm} \diamond K))| \right] \\
\le &O\left(\frac{\ln^{3/2} n}{\sqrt{n}}\right)\sqrt{\int_0^1 \mathscr{b}_K^2(x)dx} + O\left(n^{-3/4}\ln^{1/2} n\right)\sqrt{\int_0^1 K^2(x)dx}\sqrt{\int_0^1 \mathscr{b}_K^2(x)dx} \\
&+ O\left(n^{-3/2}\ln^{1/2} n\right)\int_0^1 K^2(x)dx + O\left(n^{-1}\ln^{1/2} n\right)\left(\|m\|_2 + 1\right) \\
= &O\left(\frac{\ln^{3/2} n}{\sqrt{n}}\right)\sqrt{\int_0^1 \mathscr{b}_K^2(x)dx}
\end{aligned}$$

$O(n^{-3/4}\ln^{3/2} n)\|K\|_2\|\mathscr{b}_K\|_2 = O(n^{-1/4}\ln^{3/2} n)\frac{1}{\sqrt{n}}\|K\|_2\|\mathscr{b}_K\|_2$, where both $\frac{1}{\sqrt{n}}\|K\|_2$ and $\|\mathscr{b}_K\|_2$ are smaller than $\sqrt{MISE(K)}$. And also $O(n^{-3/2}\ln^{3/2} n)\|K\|_2^2 = O(n^{-1/2}\ln^{3/2} n)MISE(K)$. Let us turn to $\frac{1}{n}(b_{\mathbb{K}} + h_m)^T\mathbb{C}\varepsilon$. As in the already considered case of $\mathbb{KC}$, different integrals of the same basis functions $\{\varphi'_{st}, \psi'_{st}\}$ are necessary for the decomposition, whereas the coefficients will be those of b_K.

$$\begin{aligned}
\frac{1}{n}(b_{\mathbb{K}} + h_m)^T\mathbb{C}\varepsilon = &\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{2}\Big(b_{\mathbb{K}}(x_{i-1}) + h_m(X_{i-1})\Big) - \Big(b_{\mathbb{K}}(x_i) + h_m(x_i)\Big) + \frac{1}{2}\Big(b_{\mathbb{K}}(x_{i+1})\right.\\
&\left. + h_m(x_{i+1})\Big)\right]\varepsilon_i \\
= &\sum_{i=1}^{n}\left[\frac{1}{2}(b_K + h_m)^{\pm}(x_{i-1}) - (b_K + h_m)^{\pm}(x_i) + \frac{1}{2}(b_K + h_m)^{\pm}(x_{i+1})\right]\varepsilon_i \\
&+ O\left(n^{-3}\right)\sqrt{\int_0^1 K^2(x)dx}\sum_{i=1}^{n}|\varepsilon_i| \\
= &\frac{1}{2}\int_{1/(2n)}^{3/(2n)}\Big(b_K(x_i + z) + h_m(x_i + z)\Big) + \Big(b_K(x_i - z) + h_m(x_i - z)\Big)dz \\
&- \int_{-1/(2n)}^{1/(2n)}\Big(b_K(x_i + z) + h_m(x_i + z)\Big)dz + O\left(n^{-3}\right)\sqrt{\int_0^1 K^2(x)dx}\sum_{i=1}^{n}|\varepsilon_i| \qquad (43)
\end{aligned}$$

We thus construct new basis vectors out of the old basis functions and use them to decompose $(b_K + h_m)^{\pm T}\mathbb{C}$.

$$\begin{aligned}\boldsymbol{\theta}'_{st} &:= \left(\int_{1/2n}^{3/2n} \varphi'_{st}(x_i+z) + \varphi'_{st}(x_i-z)dz - \int_{-1/2n}^{1/2n} \varphi'_{st}(x_i+z)dz\right)_{i=1}^{n} \\ \boldsymbol{\xi}'_{st} &:= \left(\int_{1/2n}^{3/2n} \psi'_{st}(x_i+z) + \psi'_{st}(x_i-z)dz - \int_{-1/2n}^{1/2n} \psi'_{st}(x_i+z)dz\right)_{i=1}^{n} \qquad (44)\end{aligned}$$

The set "favorable events" $B_{n3,2}$ is constructed in the usual way, so that the probability of at least one linear form of the basis functions exceeding $\lambda n^{-1/2}\ln^{1/2} n$, decreases like $n^{-\lambda^2+1/2}$. As bounds for the absolute sums of the coefficients serve those of Lemma II.2.3.

$$\begin{aligned}&\Big|\sum_{i=1}^{n}\left[\frac{1}{2}(b_K+h_m)^{\pm}(x_{i-1}) - (b_K+h_m)^{\pm}(x_i) + \frac{1}{2}(b_K+h_m)^{\pm}(x_{i+1})\right]\varepsilon_i\Big| \\ = \;& \Big|\Big[\sum_{s=1}^{s_n} \alpha'_{s,2^s-1}(b_K)\boldsymbol{\theta}'_{s,2^s-1} + \alpha'_{s_n,2^{s_n}}(b_K)\boldsymbol{\theta}'_{s_n,2^{s_n}} + \sum_{s=1}^{s_n-1}\sum_{t=1}^{2^s-1}\beta'_{st}(b_K)\boldsymbol{\xi}'_{st} \\ &+ \sum_{s=s_n}^{\max s_\kappa}\sum_{t=1}^{2^s}\beta'_{st}(b_K)\boldsymbol{\xi}'_{st}\Big]^T \varepsilon\Big| \\ = \;& O\left(\frac{\ln^{3/2} n}{\sqrt{n}}\right)\sqrt{\int_0^1 \mathscr{b}_K^2(x)dx} \qquad (45)\end{aligned}$$

The remaining term, $m^T\mathbb{C}^2\varepsilon$ again does not require a decomposition, since it does not depend on K. On the set B_{n4}, the linear form $m^T\mathbb{C}^2\varepsilon$ may be found to be bounded by $\lambda\sqrt{n\ln n}$. $P(B_{n4}^c) = O(n^{-\lambda^2})$.

$$\frac{1}{n^2}\, m^T\mathbb{C}^2\varepsilon \;\le\; \frac{\lambda\ln^{1/2} n}{n\sqrt{n}}$$

Before advancing to Lemma II.2.3, we recall equations number (20) and (23) and find

$$\begin{aligned}|\alpha'_{st}(b_K)| &= \Big|\sum_{|\kappa|\le\sqrt{n}} \widehat{b}_K(2\pi\kappa)\widehat{\varphi}'_{st}(2\pi\kappa)\Big| \\ &= \Big|\sum_{|\kappa|\le\sqrt{n}}\left(\widehat{\mathscr{b}}_K(2\pi\kappa) + O\left(n^{-5/4}\right)\sqrt{\int_0^1 K^2(x)dx}\right)\widehat{\varphi}'_{st}(2\pi\kappa)\Big| \\ &= |\alpha'_{st}(\mathscr{b}_K)| + O\left(n^{-5/4}\right)\sqrt{\int_0^1 K^2(x)dx}\sum_{|\kappa|\le\sqrt{n}}|\widehat{\varphi}'_{st}(2\pi\kappa)| \qquad (46)\end{aligned}$$

In the same way we obtain

$$\begin{aligned}|\alpha'_{st}(n\,(b_K^{\pm}\diamond K))| \;=\; & |\alpha'_{st}(\mathscr{b}_K * K)| + O\left(n^{-3/4}\right)\sqrt{\int_0^1 K^2(x)dx}\sum_{|\kappa|\le\sqrt{n}}|\widehat{\varphi}'_{st}(2\pi\kappa)| \\ &+ O\left(n^{-3/2}\right)\int_0^1 K^2(x)dx\sum_{|\kappa|\le\sqrt{n}}|\widehat{\varphi}'_{st}(2\pi\kappa)|,\end{aligned}$$

and the like for the β's.

Lemma II.2.3

$$\sum_{s=1}^{s_n} |\alpha'_{s,2^s-1}(\mathscr{b}_K)| + |\alpha'_{s_n,2^{s_n}}(\mathscr{b}_K)| \le \sqrt{\int_0^1 \mathscr{b}_K^2(x)dx}\,\sqrt{s_n+1}$$
$$\sum_{t=1}^{2^s-1} |\beta'_{st}(\mathscr{b}_K)| \le \sqrt{2\int_0^1 \mathscr{b}_K^2(x)dx} + O(n^{-1}),\ \text{for } s < s_n,$$
$$\text{and} \quad \sum_{t=1}^{2^s} |\beta'_{st}(\mathscr{b}_K)| \le 2\cdot 2^{-s/2}\|m\|_2 + O(n^{-1}),\ \text{for } s \ge s_n$$
$$\sum_{s=1}^{s_n} \sum_{|\kappa|\le\sqrt{n}} |\widehat{\varphi}'_{s,2^s-1}(2\pi\kappa)| + \sum_{|\kappa|\le\sqrt{n}} |\widehat{\varphi}'_{s_n,2^{s_n}}(2\pi\kappa)| = O(n^{1/4}\ln n)$$
$$\text{and} \quad \sum_{s=1}^{s_n-1}\sum_{t=1}^{2^s-1}\sum_{|\kappa|\le\sqrt{n}} |\widehat{\psi}'_{st}(2\pi\kappa)| + \sum_{s=s_n}^{\max s_\kappa}\sum_{t=1}^{2^s}\sum_{|\kappa|\le\sqrt{n}} |\widehat{\psi}'_{st}(2\pi\kappa)| = O(n^{1/2})$$

Exactly the same can be achieved for $\sum|\alpha'_{st}(\mathscr{b}_K * K)|$ and $\sum|\beta'_{st}(\mathscr{b}_K * K)|$ because $\widehat{\mathscr{b}_K * K} = \widehat{\mathscr{b}}_K\widehat{K}$.

Proof By the same idea as in Lemma II.2.1

$$\begin{aligned}\sum_{s=1}^{s_n} |\alpha'_{s,2^s-1}(\mathscr{b}_K)| + |\alpha'_{s_n,2^{s_n}}(\mathscr{b}_K)| &= \sum_{s=1}^{s_n}\Big|\sum_{|\kappa|\le\sqrt{n}} \widehat{\mathscr{b}}_K(2\pi\kappa)\widehat{\varphi}'_{s,2^s-1}(2\pi\kappa)\Big| \\ &\quad + \Big|\sum_{|\kappa|\le\sqrt{n}} \widehat{\mathscr{b}}_K(2\pi\kappa)\widehat{\varphi}'_{s_n,2^{s_n}}(2\pi\kappa)\Big| \\ &\le \sqrt{\int_0^1 \mathscr{b}_K^2(x)dx}\,\sqrt{s_n+1}\end{aligned}$$

Let $\sum_t$ denote again $\sum_{t=1}^{2^s-1}$ for $s \le s_n - 1$, and $\sum_{t=1}^{2^s}$ for $s \ge s_n$. A short discussion will be useful in the sequel:

$$\begin{aligned}&\sum_t \mu_{st}\,\nu_{s+1,2t-1} \sum_{|\kappa|\le\sqrt{n}} |\widehat{m}_n(2\pi\kappa)| \left[\frac{I(\kappa\ne 0)}{\kappa} + I(\kappa=0)\right] \Delta_{s+1,2t-1}(2\pi\kappa) \\ \le\ & \sum_t \mu_{st}\,\nu_{s+1,2t-1} \sum_{|\kappa|\le\sqrt{n}} |\widehat{m}_n(2\pi\kappa)|(\kappa_{s+1,2t-2}+1)^{-1}\Delta_{s+1,2t-1}(2\pi\kappa) \\ \le\ & \sqrt{\sum_t \mu_{st}^2\,\nu_{s+1,2t-1}^2(\kappa_{s+1,2t-2}+1)^{-2}\sum_{|\kappa|\le\sqrt{n}} |\widehat{m}_n(2\pi\kappa)|^2\Delta_{s+1,2t-1}(2\pi\kappa)} \\ &\quad\times\sqrt{\sum_t\sum_{|\kappa|\le\sqrt{n}} \Delta_{s+1,2t-1}(2\pi\kappa)} \\ =\ & \sqrt{\sum_t \mu_{st}^2\,\nu_{s+1,2t-1}^2(\kappa_{s+1,2t-2}+1)^{-2}\nu_{s+1,2t-1}^{-1}M_n 2^{-(s+1)}}\ \sqrt{2\sqrt{n}+1}\end{aligned}$$

$$\begin{aligned}
&\leq 2^{-(s+1)/2} M_n^{1/2} \sqrt{\sum_t (\kappa_{s+1,2t-2}+1)^{-2}}\; O(n^{1/4}) \\
&= 2^{-(s+1)/2} M_n^{1/2}\; O(n^{1/4})
\end{aligned}$$

and the same for $\nu_{s+1,2t}$ and $\Delta_{s+1,2t}$, such that

$$\begin{aligned}
&2^{(s+1)/2} M_n^{-1/2} \sum_t \mu_{st} \sum_{|\kappa|\leq\sqrt{n}} |\widehat{m}(2\pi\kappa)| \left[\frac{I(\kappa\neq 0)}{\kappa} + I(\kappa=0)\right] \\
&\times\Big[\nu_{s+1,2t-1}\Delta_{s+1,2t-1}(2\pi\kappa) + \nu_{s+1,2t}\Delta_{s+1,2t}(2\pi\kappa)\Big] \\
= \;& O(n^{1/4})
\end{aligned}$$

Now choose an arbitrary $\widetilde{\kappa}_{st} \in [\kappa_{s,t-1}, \kappa_{st}]$.

$$\begin{aligned}
&\sum_t |\beta'_{st}(\mathscr{b}_K)| \\
=\;& \sum_t \Big| \sum_{|\kappa|\leq\sqrt{n}} \widehat{\mathscr{b}}_K(2\pi\kappa)\overline{\widehat{\psi'}}_{st}(2\pi\kappa)\Big| \\
=\;& \sum_t \Big| \sum_{|\kappa|\leq\sqrt{n}} \widehat{m}(2\pi\kappa)\left(1-\widehat{K}(2\pi\kappa)\right)\mu_{st}\; 2^{(s+1)/2} M_n^{-1/2}\overline{\widehat{m}}_n(2\pi\kappa)\Big[\nu_{s+1,2t-1}\Delta_{s+1,2t-1}(2\pi\kappa) \\
&-\nu_{s+1,2t}\Delta_{s+1,2t}(2\pi\kappa)\Big]\Big| \\
\leq\;& \sum_t \Big| \sum_{|\kappa|\leq\sqrt{n}} |\widehat{m}_n(2\pi\kappa)|^2 \left(1-\widehat{K}(2\pi\kappa)\right)\mu_{st}\; 2^{(s+1)/2} M_n^{-1/2}\Big[\nu_{s+1,2t-1}\Delta_{s+1,2t-1}(2\pi\kappa) \\
&-\nu_{s+1,2t}\Delta_{s+1,2t}(2\pi\kappa)\Big]\Big| + \sum_t \Big| \sum_{|\kappa|\leq\sqrt{n}} n^{-5/4}\left[\frac{I(\kappa\neq 0)}{\kappa} + I(\kappa=0)\right] |\widehat{m}_n(2\pi\kappa)| \\
&\times\mu_{st}\; 2^{(s+1)/2} M_n^{-1/2}\Big[\nu_{s+1,2t-1}\Delta_{s+1,2t-1}(2\pi\kappa) + \nu_{s+1,2t}\Delta_{s+1,2t}(2\pi\kappa)\Big]\Big| \\
\leq\;& \sum_t \mu_{st}\; 2^{(s+1)/2} M_n^{-1/2}\Big| \sum_{|\kappa|\leq\sqrt{n}} |\widehat{m}_n(2\pi\kappa)|^2 \left(1-\widehat{K}(2\pi\widetilde{\kappa}_{st})\right)\Big[\nu_{s+1,2t-1}\Delta_{s+1,2t-1}(2\pi\kappa) \\
&-\nu_{s+1,2t}\Delta_{s+1,2t}(2\pi\kappa)\Big] + \sum_{|\kappa|\leq\sqrt{n}} |\widehat{m}_n(2\pi\kappa)|^2\left(\widehat{K}(2\pi\widetilde{\kappa}_{st}) - \widehat{K}(2\pi\kappa)\right) \\
&\times\Big[\nu_{s+1,2t-1}\Delta_{s+1,2t-1}(2\pi\kappa) - \nu_{s+1,2t}\Delta_{s+1,2t}(2\pi\kappa)\Big]\Big| + O(n^{-1}) \\
=\;& \sum_t \mu_{st}\; 2^{(s+1)/2} M_n^{-1/2}\Big| \left(1-\widehat{K}(2\pi\widetilde{\kappa}_{st})\right)\Big[\nu_{s+1,2t-1}2^{-s-1}M_n\nu^{-1}_{s+1,2t-1}(2\pi\kappa) \\
&-\nu_{s+1,2t}2^{-s-1}M_n\nu^{-1}_{s+1,2t}(2\pi\kappa)\Big] + \sum_{|\kappa|\leq\sqrt{n}} |\widehat{m}_n(2\pi\kappa)|^2\left(\widehat{K}(2\pi\widetilde{\kappa}_{st}) - \widehat{K}(2\pi\kappa)\right) \\
&\times\Big[\nu_{s+1,2t-1}\Delta_{s+1,2t-1}(2\pi\kappa) - \nu_{s+1,2t}\Delta_{s+1,2t}(2\pi\kappa)\Big]\Big| + O(n^{-1}) \\
=\;& \sum_t \mu_{st}\; 2^{(s+1)/2} M_n^{-1/2}\;\Big|\; 0 + \sum_{|\kappa|\leq\sqrt{n}} |\widehat{m}_n(2\pi\kappa)|^2\left(\widehat{K}(2\pi\widetilde{\kappa}_{st}) - \widehat{K}(2\pi\kappa)\right) \\
&\times\Big[\nu_{s+1,2t-1}\Delta_{s+1,2t-1}(2\pi\kappa) - \nu_{s+1,2t}\Delta_{s+1,2t}(2\pi\kappa)\Big]\Big| + O(n^{-1})
\end{aligned}$$

$$\begin{aligned}
&\leq \sum_t \mu_{st} 2^{(s+1)/2} M_n^{-1/2} \sum_{|\kappa|\leq\sqrt{n}} |\widehat{m}_n(2\pi\kappa)|^2 \left(\widehat{K}(2\pi\kappa_{s,t-1}+1) - \widehat{K}(2\pi\kappa_{st})\right) \\
&\qquad \times\left[\nu_{s+1,2t-1}\Delta_{s+1,2t-1}(2\pi\kappa) + \nu_{s+1,2t}\Delta_{s+1,2t}(2\pi\kappa)\right] + O(n^{-1}) \\
&= \sum_t \mu_{st}\ 2^{(s+1)/2} M_n^{1/2} \left(\widehat{K}(2\pi\kappa_{s,t-1}+1) - \widehat{K}(2\pi\kappa_{st})\right) \cdot 2\cdot 2^{-(s+1)} + O(n^{-1}) \\
&= 2\cdot 2^{-(s+1)/2} M_n^{1/2} \sum_t \mu_{st} \left(\widehat{K}(2\pi\kappa_{s,t-1}+1) - \widehat{K}(2\pi\kappa_{st})\right) + O(n^{-1})
\end{aligned}$$

$\mu_{st} \leq 2^{-1/2}$, and for $s \geq s_n$ we can replace $\sum \widehat{K}(2\pi\kappa_{s,t-1}+1) - \widehat{K}(2\pi\kappa_{st})$ by $\widehat{K}(0) - \widehat{K}(2\pi\lfloor\sqrt{n}\rfloor) \leq 1$, so that

$$\begin{aligned}
\sum_{t=1}^{2^s} |\beta'_{st}(\mathscr{b}_K)| &\leq \sqrt{2}\cdot 2^{-(s+1)/2} M_n^{1/2} + O(n^{-1}) \\
&\leq 2^{-s/2}\left(2\|m\|_2 + O(n^{-5/4})\right) + O(n^{-1}) \\
&= 2\cdot 2^{-s/2}\|m\|_2 + O(n^{-1})
\end{aligned}$$

Otherwise for $s < s_n$:

$$\begin{aligned}
\sum_{t=1}^{2^s-1} |\beta'_{st}(\mathscr{b}_K)| &\leq \sqrt{2}\cdot 2^{-(s+1)/2} M_n^{1/2} \left(\widehat{K}(0) - \widehat{K}(2\pi\kappa_{s,2^s-1})\right) + O(n^{-1}) \\
&= 2^{-s/2} M_n^{1/2} \left(1 - \widehat{K}(2\pi\kappa_{s,2^s-1})\right) + O(n^{-1}) \\
&= 2^{-s/2} M_n^{1/2} \left[\sum_{|\kappa|\leq\sqrt{n}} \widehat{\varphi}'_{s,2^s}(2\pi\kappa)^2\right] \left(1 - \widehat{K}(2\pi\kappa_{s,2^s-1})\right) + O(n^{-1}) \\
&= \nu_{s,2^s}\ 2^{s/2} M_n^{-1/2} \sum_{|\kappa|\leq\sqrt{n}} |\widehat{m}_n(2\pi\kappa)|^2 \Delta_{s,2^s}(2\pi\kappa) \left(1 - \widehat{K}(2\pi\kappa_{s,2^s-1})\right) + O(n^{-1}) \\
&\leq \nu_{s,2^s}\ 2^{s/2} M_n^{-1/2} \sum_{|\kappa|\leq\sqrt{n}} |\widehat{m}_n(2\pi\kappa)|^2 \Delta_{s,2^s}(2\pi\kappa) \left(1 - \widehat{K}(2\pi\kappa)\right) + O(n^{-1}) \\
&= \nu_{s,2^s}^{1/2} \sum_{|\kappa|\leq\sqrt{n}} \widehat{m}_n(2\pi\kappa) \left(1 - \widehat{K}(2\pi\kappa)\right) \overline{\widehat{\varphi}'}_{s,2^s}(2\pi\kappa) + O(n^{-1}) \\
&= 2\nu_{s,2^s}^{1/2} \sum_{|\kappa|\leq\sqrt{n}} \left[\widehat{\mathscr{b}}_K(2\pi\kappa) + n^{-5/4}\left[\frac{I(\kappa\neq 0)}{\kappa} + I(\kappa=0)\right] \left(1 - \widehat{K}(2\pi\kappa)\right)\right] \\
&\qquad \times \overline{\widehat{\varphi}'}_{s,2^s}(2\pi\kappa) + O(n^{-1}) \\
&\leq \nu_{s,2^s}^{1/2} \left[\sqrt{\sum_{|\kappa|\leq\sqrt{n}} |\widehat{\mathscr{b}}_K(2\pi\kappa)|^2} + n^{-5/4}\sqrt{\sum_{|\kappa|\leq\sqrt{n}} \left[\frac{I(\kappa\neq 0)}{\kappa} + I(\kappa=0)\right]^2}\right] \\
&\qquad \times \sqrt{\sum_{|\kappa|\leq\sqrt{n}} |\widehat{\varphi}'_{s,2^s}(2\pi\kappa)|^2} + O(n^{-1}) \\
&\leq \nu_{s,2^s}^{1/2} \left[\sqrt{\int_0^1 \mathscr{b}_K^2(x)dx} + \sqrt{5}\, n^{-5/4}\right] + O(n^{-1})
\end{aligned}$$

$$= \sqrt{2}\sqrt{\int_0^1 \ell_K^2(x)dx} + O(n^{-1})$$

With $\widehat{(\ell_K * K)}(\kappa) = \widehat{\ell}_K(\kappa)\widehat{K}(\kappa)$ the same procedure may be applied for $\sum|\alpha'_{st}(\ell_K * K)|$ and $\sum|\beta'_{st}(\ell_K * K)|$. Still we have to consider the residual terms which result from the approximations in Subsection 2.1. Because of orthonormality:

$$\begin{aligned}
\sum_{|\kappa|\le\sqrt{n}}\left[\sum_{s=1}^{s_n}|\widehat{\varphi}'_{s,2^s-1}(2\pi\kappa)| + |\widehat{\varphi}'_{s_n,2^{s_n}}(2\pi\kappa)|\right] &\le \sqrt{\sum_{s=1}^{s_n}\int_0^1 \varphi'_{s,2^s-1}(x)^2dx + \int_0^1 \varphi'_{s_n,2^{s_n}}(x)^2dx} \\
&\quad \times \sqrt{\sum_{|\kappa|\le\sqrt{n}} 1} \\
&\le O(n^{1/4}\ln n)
\end{aligned}$$

We remember that s_κ was chosen so as to equal $\min\{s \in \mathbb{N} | 2^{-s}M_n \le 2|\widehat{m}_n(2\pi\kappa)|^2\}$. It follows that $|\widehat{m}_n(2\pi\kappa)| < 2^{-s_\kappa/2}M_n^{1/2}$.

$$\begin{aligned}
\sum_{s=1}^{\max s_\kappa}\sum_{t=1}^{2^s}\sum_{|\kappa|\le\sqrt{n}}|\widehat{\psi}'_{st}(2\pi\kappa)| &= \sum_{|\kappa|\le\sqrt{n}}\sum_{s=1}^{s_\kappa}\sum_{t=1}^{2^s} 2^{(s+1)/2}\mu_{st}M_n^{-1/2}|\widehat{m}_n(2\pi\kappa)|\Big[\nu_{s+1,2t-1}\Delta_{s+1,2t-1}(2\pi\kappa) \\
&\qquad + \nu_{s+1,2t}\Delta_{s+1,2t}(2\pi\kappa)\Big] \\
&\le \sqrt{2}\ M_n^{-1/2}\sum_{|\kappa|\le\sqrt{n}}|\widehat{m}_n(2\pi\kappa)|\sum_{s=1}^{s_\kappa}2^{(s+1)/2}\sum_{t=1}^{2^s}\Big[\Delta_{s+1,2t-1}(2\pi\kappa) \\
&\qquad + \Delta_{s+1,2t}(2\pi\kappa)\Big] \\
&= \sqrt{2}\ M_n^{-1/2}\sum_{|\kappa|\le\sqrt{n}}|\widehat{m}_n(2\pi\kappa)|\sum_{s=1}^{s_\kappa}2^{(s+1)/2}\sum_{t=1}^{2^{s+1}}\Delta_{s+1,t}(2\pi\kappa) \\
&= \sqrt{2}\ M_n^{-1/2}\sum_{|\kappa|\le\sqrt{n}}|\widehat{m}_n(2\pi\kappa)|\sum_{s=1}^{s_\kappa}2^{(s+1)/2}\cdot 1 \\
&< \sqrt{2}\sum_{|\kappa|\le\sqrt{n}}2^{-s_\kappa/2}\sum_{s=1}^{s_\kappa}2^{(s+1)/2} \\
&< \frac{\sqrt{2}}{\sqrt{2}-1}\sum_{|\kappa|\le\sqrt{n}}2^{-s_\kappa/2}\ 2^{(s_\kappa+1)/2} \\
&= O(n^{1/2}) \qquad \square
\end{aligned}$$

Lemma II.2.4

$$P\left(|\boldsymbol{\varphi}'^{\,T}_{st}\varepsilon| > \frac{\lambda\ln^{1/2}n}{\sqrt{n}}\right) = O\left(n^{-\lambda^2}\right) \quad \text{and} \quad P\left(|\boldsymbol{\psi}'^{\,T}_{st}\varepsilon| > \frac{\lambda\ln^{1/2}n}{\sqrt{n}}\right) = O\left(n^{-\lambda^2}\right) \quad \forall\ s,t$$

and the like for $\boldsymbol{\theta}'_{st}$ and $\boldsymbol{\xi}'_{st}$. Furthermore:

$$P\left(|m^T\mathbb{C}^2\varepsilon| > \lambda\sqrt{n}\ln^{1/2}n\right) = O\left(n^{-\lambda^2}\right)$$

Proof We use once more Bentkus/Rudzkis' exponential inequality for polynomial forms (see Appendix A.3):

$$P\left(|a^T\varepsilon| > x\right) \quad \leq \quad 2\exp\left\{-\frac{x^2}{2\|a\|_2^2}\right\}$$

Substituting a by $\boldsymbol{\varphi}'_{st}$,

$$\begin{aligned}\|\boldsymbol{\varphi}'_{st}\|_2^2 &= \sum_{i=1}^{n}\left(\int_{-1/2n}^{1/2n}\varphi'_{st}(x_i+y)dy\right)^2 \\ &\leq \sum_{i=1}^{n}\frac{1}{n}\int_{-1/2n}^{1/2n}{\varphi'_{st}}^2(x_i+y)dy \\ &= \frac{1}{n}\int_0^1{\varphi'_{st}}^2(y)dy \\ &= \frac{1}{n}\end{aligned}$$

Hence it holds that

$$\begin{aligned}P\left(|{\boldsymbol{\varphi}'_{st}}^T\varepsilon| > \frac{\lambda\ln^{1/2}n}{\sqrt{n}}\right) &\leq 2\exp\left\{-\frac{\frac{\lambda^2\ln n}{n}}{\frac{2}{n}}\right\} \\ &= 2\exp\left\{-\frac{\lambda^2\ln n}{2}\right\} \\ &= O\Big(\exp\left\{-\lambda^2\ln n\right\}\Big)\end{aligned}$$

For the vectors $\boldsymbol{\psi}'_{st}$, $\boldsymbol{\theta}'_{st}$ and $\boldsymbol{\xi}'_{st}$, the same calculation applies.
Since $\|m^T\mathbb{C}^2\|_2^2 \leq 16n(\int_0^1 m^2(x)dx + \frac{1}{n^2}\max|m''|)$, the appropriate bound for $m^T\mathbb{C}^2\varepsilon$ can only be $\lambda\sqrt{n}\ln^{1/2}n$. □

To approximate the probability of the union of all complementary sets belonging to the $\boldsymbol{\psi}'$s, we have to regard the number of mother wavelet functions, which are necessary for the decomposition of $\widehat{\boldsymbol{b}}_K$. Instead of counting the functions ψ_{st} over the indices s and t, we can roughly bound their number by the sum of the $s_\kappa = \min\{s\in\mathbb{N}|2^{-s}M_n \leq 2|\widehat{m}_n(2\pi\kappa)|^2\}$ over κ. $2^{-s_\kappa} > M_n^{-1}|\widehat{m}_n(2\pi\kappa)|^2$, so that

$$\begin{aligned}\sum_{|\kappa|\leq\sqrt{n}} s_\kappa &< \frac{2\sqrt{n}+1}{\ln 2}\ln M_n - \frac{2}{\ln 2}\sum_{|\kappa|\leq\sqrt{n}}\ln|\widehat{m}_n(2\pi\kappa)| \\ &< \frac{2\sqrt{n}+1}{\ln 2}\ln M_n - \frac{2}{\ln 2}\sum_{|\kappa|\leq\sqrt{n}}\ln\left|n^{-5/4}\left[\frac{I(\kappa\neq 0)}{\kappa}+I(\kappa=0)\right]\right| \\ &= O\left(\sqrt{n}\ln n\right)\end{aligned}$$

Together with the s_n+1 father wavelets we have

$$P\left(B^c_{n3,1}\right) \quad\leq\quad \sum_{s=1}^{s_n} P\left(|{\boldsymbol{\varphi}'_{s,2^s-1}}^T\varepsilon| > \frac{\lambda\ln n}{\sqrt{n}}\right) + P\left(|{\boldsymbol{\varphi}'_{s_n,2^{s_n}}}^T\varepsilon| > \frac{\lambda\ln n}{\sqrt{n}}\right)$$

$$
\begin{aligned}
& + \sum_{s=1}^{\max s_\kappa} \sum_{t=1}^{2^s} P\left(|\boldsymbol{\psi}'^{T}_{st}\varepsilon| > \frac{\lambda \ln n}{\sqrt{n}} \right) \\
= \;& O(\ln n)\, O\left(n^{-\lambda^2}\right) + O\left(\sqrt{n} \ln n\right) O\left(n^{-\lambda^2}\right) \\
= \;& O\left(n^{-\lambda^2+1/2}\right) \qquad (47)
\end{aligned}
$$

The same is of course true for the $\boldsymbol{\theta}'$s and $\boldsymbol{\xi}'$s.
Finally, we have to answer the question if also for $\sum |\varepsilon_i|$ there holds a sufficiently strong inequality. We may easily check that for $|\varepsilon_i|$ with $\varepsilon_i \sim \mathcal{N}(0,1)$, $E|\varepsilon_i| = \sqrt{2/\pi}$, $\sigma^2 = Var|\varepsilon_i| = 1 - 2/\pi$ and the absolute central moments $E|\,|\varepsilon_i| - E|\varepsilon_i|\,|^k$ are bounded by $2k!\,\sigma^2$, such that Bernstein's condition is fulfilled and for $0 < \tau \le n\sigma^2/4$, it holds that

$$
P\left(\left| \sum_{i=1}^{n} |\varepsilon_i| - E|\varepsilon_i| \right| \ge \tau \right) \le 2 \exp\left\{ -\frac{\tau^2}{4n\sigma^2} \right\} \qquad (48)
$$

(Petrov (1987)). Choose $\tau = n\sigma^2/4$, so there results a comfortable probability of exceedance of $O(e^{-n})$.

3 E[CV]-MISE

Let us now consider the non-random part of CV-ISE. Except for an additional σ^2, CV is an asymptotically unbiased estimator for MISE. However, since σ^2 does not depend on K, it drops out when subtracting $E[CV(K_0)] - MISE(K_0) + MISE(K^*) - E[CV(K^*)]$. What we have to verify is whether $E[CV(K)] - \sigma^2 - MISE(K)$ is negligible in proportion to MISE.
Within $E[CV(K)] - MISE(K)$ we can distinguish a bias stemming from the discretization and a bias attributable to the "leave-one-out" method. The former will obviously essentially depend on the smoothness properties of K and m.

$$
\begin{aligned}
& E[CV(K)] \\
= \;& E\left[\frac{1}{n} Y^T \left(\mathbb{K} - I + K^{\pm}(0)\mathbb{C} \right)^2 Y \right] \\
= \;& \frac{1}{n} m^T \left(\mathbb{K} - I + K^{\pm}(0)\mathbb{C} \right)^2 m + \frac{1}{n} E\left[\varepsilon^T \left(\mathbb{K} - I + K^{\pm}(0)\mathbb{C} \right)^2 \varepsilon \right] \\
= \;& \frac{1}{n} b_{\mathbb{K}}^T b_{\mathbb{K}} + \frac{2}{n} K^{\pm}(0) b_{\mathbb{K}}^T \mathbb{C} m + \frac{1}{n} K^{\pm}(0)^2 m^T \mathbb{C}^2 m + \frac{1}{n} E\left[\varepsilon^T (\mathbb{K} - I)^2 \varepsilon \right] \\
& + \frac{1}{n} E\left[\varepsilon^T \left(2K^{\pm}(0)\mathbb{K}\mathbb{C} - 2K^{\pm}(0)\mathbb{C} + K^{\pm}(0)^2 \mathbb{C}^2 \right) \varepsilon \right] \\
= \;& \frac{1}{n} \sum_{i=1}^{n} b_{\mathbb{K}}^2(x_i) + \frac{2}{n} K^{\pm}(0) \sum_{i=1}^{n} b_{\mathbb{K}}(x_i) \left(\frac{1}{2} m(x_{i-1}) - m(x_i) + \frac{1}{2} m(x_{i+1}) \right) \\
& + \frac{1}{n} K^{\pm}(0)^2 \sum_{i=1}^{n} m(x_i) \left(\frac{1}{4} m(x_{i-2}) - m(x_{i-1}) + \frac{3}{2} m(x_i) - m(x_{i+1}) + \frac{1}{4} m(x_{i+2}) \right) \\
& + \frac{1}{n} \sigma^2 \sum_{i,j} K^{\pm}(x_i - x_j)^2 + \sigma^2 \left(1 + 2K^{\pm}(0) K^{\pm}(1/n) - \frac{1}{2} K^{\pm}(0)^2 \right)
\end{aligned}
$$

MISE has already been worked out in Section 2, equation (9).

$$
MISE(K) = \int_0^1 b_{\mathbb{K}}^2(x)\,dx + \sigma^2 \sum_{i=1}^{n} \int_0^1 K^{\pm}(x_i - x)^2\,dx
$$

$$= \left(\int_0^1 \mathcal{b}_K^2(x)dx + \frac{\sigma^2}{n}\int_0^1 K^2(x)dx\right)\left(1+O(n^{-1/2})\right)$$

$\frac{1}{n}b_{\mathbb{K}}^T b_{\mathbb{K}}$ is the discrete version of $\int b_{\mathbb{K}}^2(x)dx$, and $\frac{1}{n}\sum K^{\pm 2}(x_i - x_j)$ the discrete version of $\sum\int K^{\pm 2}(x_i - x)dx$. For $n \longrightarrow \infty$ the sums obviously tend to the integrals. However, it remains to be checked whether the errors vanish rapidly enough. For $b_{\mathbb{K}}$ we split the difference into 3 summands:

$$\begin{aligned}\left|\frac{1}{n}\sum_{i=1}^n b_{\mathbb{K}}^2(x_i) - \int_0^1 b_{\mathbb{K}}^2(x)dx\right| \quad &\leq \quad \left|\frac{1}{n}\sum_{i=1}^n b_{\mathbb{K}}^2(x_i) - \frac{1}{n}\sum_{i=1}^n \mathcal{b}_{\mathbb{K}}^2(x_i)\right| + \left|\frac{1}{n}\sum_{i=1}^n \mathcal{b}_{\mathbb{K}}^2(x_i) - \int_0^1 \mathcal{b}_{\mathbb{K}}^2(x)dx\right| \\ &\quad + \left|\int_0^1 \mathcal{b}_{\mathbb{K}}^2(x)dx - \int_0^1 b_{\mathbb{K}}^2(x)dx\right| \qquad (49)\end{aligned}$$

We already know that the last summand is $O(n^{-1/2})MISE(K)$ (inequality (9)), so it will not pose any problems. For the second one we revert to (7):

$$|b_{\mathbb{K}}(x) - \mathcal{b}_K(x)| \quad \leq \quad O\left(n^{-1}\right)\sqrt{\int_0^1 K^2(x)dx}$$

where, of course the $O(.)$ depends on m. For all $j < \infty$:

$$\begin{aligned}|\mathcal{b}_K^{(j)}(x)| \quad &= \quad \left|(K*m)^{(j)}(x) - m^{(j)}(x)\right| \quad = \quad \left|-(m^{(j)}*K)(x) - m^{(j)}(x)\right| \\ &\leq \quad \left(\sqrt{\int_0^1 K^2(x)dx} + 1\right)\max|m^{(j)}|\end{aligned}$$

We plug this formula into

$$\begin{aligned}&\left|\frac{1}{n}\sum_{i=1}^n b_{\mathbb{K}}^2(x_i) - \frac{1}{n}\sum_{i=1}^n \mathcal{b}_{\mathbb{K}}^2(x_i)\right| \\ = \quad &\left|\frac{1}{n}\sum_{i=1}^n \left(b_{\mathbb{K}}(x_i) - \mathcal{b}_{\mathbb{K}}(x_i)\right)\left(2\mathcal{b}_K(x_i) + b_{\mathbb{K}}(x_i) - \mathcal{b}_{\mathbb{K}}(x_i)\right)\right| \\ = \quad &O\left(n^{-1}\right)\sqrt{\int_0^1 K^2(x)dx}\;\frac{1}{n}\sum_{i=1}^n\left(2|\mathcal{b}_K(x_i)| + O\left(n^{-1}\right)\sqrt{\int_0^1 K^2(x)dx}\right) \\ = \quad &O\left(n^{-1}\right)\sqrt{\int_0^1 K^2(x)dx}\;\left(2\sqrt{\int_0^1 \mathcal{b}_K^2(x)dx} + \frac{2}{n^2}\max|\mathcal{b}_K''| + O\left(n^{-1}\right)\sqrt{\int_0^1 K^2(x)dx}\right) \\ = \quad &O\left(n^{-1}\right)\sqrt{\int_0^1 K^2(x)dx}\;\left(\sqrt{\int_0^1 \mathcal{b}_K^2(x)dx} + O\left(n^{-2}\right)\left(\sqrt{\int_0^1 K^2(x)dx} + 1\right)\right. \\ &\qquad \left. + O\left(n^{-1}\right)\sqrt{\int_0^1 K^2(x)dx}\right) \\ = \quad &O\left(n^{-1/2}\right)MISE(K)\end{aligned}$$

Together with the remaining summand, which is bounded next, we have for (47): $|\frac{1}{n}\sum b_{\mathbb{K}}^2(x_i) - \int_0^1 b_{\mathbb{K}}^2(x)dx| = O(n^{-1/2})MISE(K)$.

$$\begin{aligned}
\Big|\frac{1}{n}\sum_{i=1}^{n} b_{\mathbb{K}}^2(x_i) - \int_0^1 b_{\mathbb{K}}^2(x)dx\Big| &\leq \frac{1}{n^2}\max|(b_K^2)''| \\
&= \frac{2}{n^2}\max|b_K \cdot b_K'' + (b_K')^2| \\
&\leq O\left(n^{-2}\right)\left(\sqrt{\int_0^1 K^2(x)dx} + 1\right)^2 \\
&= O\left(n^{-1}\right) MISE(K)
\end{aligned}$$

In the calculation for the variance part we have to make use of the assumption $\widehat{K}(2\pi\kappa) = 0$ for all $|\kappa| > \sqrt{n}$, which induces $\max|K| \leq \|\widehat{K}\|_1 = \sum|\widehat{K}(2\pi\kappa)| \leq 2n^{1/2} + 1$, $\max|K'| \leq 4\pi n$ and $\max|K''| \leq 8\pi^2 n^{3/2}$, such that we have:

$$\begin{aligned}
\frac{1}{n}\sum_{i,j} K^{\pm 2}(x_i - x_j) - \sum_{i=1}^{n}\int_0^1 K^{\pm 2}(x_i - x)dx &\leq \frac{1}{n^2}\max\Big|\left((K^{\pm})^2\right)''\Big| \\
&\leq \frac{2}{n^2}\max\Big|K^{\pm}\cdot(K^{\pm})'' + \left((K^{\pm})'\right)^2\Big| \\
&\leq \frac{2}{n^2}\left(\frac{1}{n}\max|K| \cdot \frac{1}{n}\max|K''| + \frac{1}{n^2}\max{}^2|K'|\right) \\
&= O(n^{-2})
\end{aligned}$$

Hence the error stemming from the discretization is negligible in proportion to MISE. And so is the error which is due to cross-validation, since $K^{\pm}(0) \leq (2\sqrt{n}+1)/n$ and also $K^{\pm}(1/n) \leq (2\sqrt{n}+1)/n$. With similar considerations as before we can yield:

$$\begin{aligned}
&\frac{2}{n}\, K^{\pm}(0)\sum_{i=1}^{n} b_{\mathbb{K}}(x_i)\left(\frac{1}{2}m(x_{i-1}) - m(x_i) + \frac{1}{2}m(x_{i+1})\right) \\
=\ & O\left(n^{-1/2}\right)\left(\sqrt{\int_0^1 b_K^2(x)dx} + O\left(n^{-1}\right)\sqrt{\int_0^1 K^2(x)dx} + O\left(n^{-2}\right)\left(\sqrt{\int_0^1 K^2(x)dx} + 1\right)\right)
\end{aligned}$$

$$\begin{aligned}
&\frac{1}{n}\, K^{\pm}(0)^2\sum_{i=1}^{n} m(x_i)\left(\frac{1}{4}m(x_{i-2}) - m(x_{i-1}) + \frac{3}{2}m(x_i) - m(x_{i+1}) + \frac{1}{4}m(x_{i+2})\right) \\
=\ & O\left(n^{-1}\right)
\end{aligned}$$

$$\begin{aligned}
&\sigma^2\left(2K^{\pm}(0)K^{\pm}(1/n) + \frac{1}{2}K^{\pm}(0)^2\right) \\
=\ & \sigma^2 O(n^{-1})
\end{aligned}$$

4 ISE-MISE

This section follows the same lines as the determination of the bound for CV-E[CV]. It will be seen that the difference ISE-MISE comprises a quadratic and a linear form, which will be decomposed by means of the two wavelet basis defined in Section 2 in order to find bounds.

$$\begin{aligned} ISE(K) &= \int_0^1 \Big(\widetilde{m}_K(x) - m(x)\Big)^2 dx \\ &= \int_0^1 \left(\sum_{i=1}^n K^{\pm}(x_i - x)m(x_i) - m(x) + \sum_{i=1}^n K^{\pm}(x_i - x)\varepsilon_i\right)^2 dx \\ &= \int_0^1 b_{\mathbb{K}}^2(x)dx + 2\int_0^1 b_{\mathbb{K}}(x)\left(\sum_{i=1}^n K^{\pm}(x_i - x)\varepsilon_i\right)dx + \int_0^1 \left(\sum_{i=1}^n K^{\pm}(x_i - x)\varepsilon_i\right)^2 dx \\ &= \int_0^1 b_{\mathbb{K}}^2(x)dx + 2\sum_{j=1}^n \int_0^1 b_{\mathbb{K}}(x)K^{\pm}(x_j - x)dx\ \varepsilon_j + \sum_{i,j}\int_0^1 K^{\pm}(x_i - x)K^{\pm}(x_j - x)dx\ \varepsilon_i\varepsilon_j \end{aligned} \tag{50}$$

$$\begin{aligned} ISE(K) - MISE(K) &= 2\sum_{j=1}^n \int_0^1 b_{\mathbb{K}}(x)K^{\pm}(x_j - x)dx\ \varepsilon_j + \sum_{i,j}\varepsilon_i\varepsilon_j \int_0^1 K^{\pm}(x_i - x)K^{\pm}(x_j - x)dx \\ &\quad -E\left[\sum_{i,j}\varepsilon_i\varepsilon_j \int_0^1 K^{\pm}(x_i - x)K^{\pm}(x_j - x)dx\right] \end{aligned} \tag{51}$$

Since the pursuit of a bound for the wavelet coefficients will take place in the Fourier domain, we will start by calculating the Fourier coefficients of the integrals appearing as weights in the polynomial forms. By means that we already abundantly used in Subsection 2.1, we derive:

$$\int_0^1 K^{\pm}(x_i - x)K^{\pm}(x_j - x)dx = \left(K^{\pm} * K\right)^{\pm}(x_i - x_j) \tag{52}$$

$$\begin{aligned} (\widehat{K^{\pm} * K})(2\pi\kappa) &= \widehat{K^{\pm}}(2\pi\kappa)\widehat{K}(2\pi\kappa) \\ &= \int_0^1 K^{\pm}(x)\ e^{i2\pi\kappa x}dx\ \widehat{K}(2\pi\kappa) \\ &= \int_0^1 \int_{-1/2n}^{1/2n} K(x+y)dy\ e^{i2\pi\kappa x}dx\ \widehat{K}(2\pi\kappa) \\ &= \int_{-1/2n}^{1/2n}\int_0^1 K(x+y)e^{i2\pi\kappa(x+y)}dx\, e^{-i2\pi\kappa y}dy\ \widehat{K}(2\pi\kappa) \\ &= \int_{-1/2n}^{1/2n} e^{-i2\pi\kappa y}dy\ \widehat{K}^2(2\pi\kappa) \\ &= \frac{1}{\pi\kappa}\sin\left(\frac{\pi\kappa}{n}\right)\ \widehat{K}^2(2\pi\kappa) \\ &= \left(\frac{1}{n} + O\left(n^{-2}\right)\right)\widehat{K}^2(\kappa) \end{aligned} \tag{53}$$

and

$$\begin{aligned}\int_0^1 b_{\mathbb{K}}(y)K^{\pm}(x-y)dy &= \int_0^1 n\Big(b_K^{\pm}(y)+O\left(n^{-3}\right)\Big)K^{\pm}(x-y)dy \\ &= n\int_0^1 b_K^{\pm}(y)K^{\pm}(x-y)dy + O\left(n^{-3}\right)\sqrt{\int_0^1 K^2(x)dx} \\ &= n\left(K*b_K^{\pm}\right)^{\pm}(x) + O\left(n^{-3}\right)\sqrt{\int_0^1 K^2(x)dx} \qquad (54)\end{aligned}$$

The required Fourier coefficients are obtained in the same way as for $(\widehat{K^{\pm}*K})(2\pi\kappa)$.

$$\begin{aligned} n\big(\widehat{K*b_K^{\pm}}\big)(2\pi\kappa) &= n\,\widehat{b_K^{\pm}}(2\pi\kappa)\widehat{K}(2\pi\kappa) \\ &= n\left(\frac{1}{n}+O\left(n^{-2}\right)\right)\widehat{b}_K(2\pi\kappa)\widehat{K}(2\pi\kappa) \\ &= \Big(1+O\left(n^{-1}\right)\Big)\left(\widehat{\mathscr{b}}_K(2\pi\kappa)+O\left(n^{-5/4}\right)\sqrt{\int_0^1 K^2(x)dx}\right)\widehat{K}(2\pi\kappa) \qquad (55)\end{aligned}$$

With $\widehat{K}(2\pi\kappa)\le 1$, $|\widehat{\mathscr{b}}_K(2\pi\kappa)|\le\int_0^1|\mathscr{b}_K(x)|dx\le\|\mathscr{b}_K\|_2$, we have:

$$\left|n\big(\widehat{K*b_K^{\pm}}\big)(2\pi\kappa)-\widehat{\mathscr{b}}_K(2\pi\kappa)\widehat{K}(2\pi\kappa)\right| = O\left(n^{-5/4}\right)\sqrt{\int_0^1 K^2(x)dx}+O\left(n^{-1}\right)\sqrt{\int_0^1 \mathscr{b}_K^2(x)dx} \quad (56)$$

Again we decompose the quadratic form into a the stochastic part, which is bounded on the set of "favorable events" $B_{n1,1}$ by $\lambda\ln n$ as in Lemma II.2.2, and the sum of absolute wavelet coefficients, that is in turn of order $O(n^{-1}\ln n)\|K\|_2+O(n^{-3/2})\|K\|_2^2$ like in Lemma II.2.1. For the linear form it is this time not necessary to suppress the high-frequency component, since multiplication of $\widehat{\mathscr{b}}_K$ by $\widehat{K}$ already does it. We proceed like in Lemmata II.2.3 and II.2.4 and get the bound $\lambda n^{-1/2}\ln^{1/2}n$ for the stochastic part on $B_{n3,1}$ and for the coefficients $O(\ln n)\|\mathscr{b}_K\|_2+O(n^{-3/4})\|K\|_2^2+O(n^{-1/2})\|K\|_2$, whereas the remainder term $O(n^{-3})\|K\|_2 \cdot\sum|\varepsilon_i|$ is of order $O(n^{-2})\ \|K\|_2$ on $B_{n3,0}$. So evaluating ISE-MISE on $B_{n1,1}\cap B_{n3,1}\cap B_{n3,0}$:

$$\begin{aligned} & ISE(K)-MISE(K) \\ =\ & \sum_{i,j}\varepsilon_i\varepsilon_j\int_0^1 K^{\pm}(x_i-x)K^{\pm}(x_j-x)dx - E\left[\sum_{i,j}\varepsilon_i\varepsilon_j\int_0^1 K^{\pm}(x_i-x)K^{\pm}(x_j-x)dx\right] \\ & +2\sum_{j=1}^n\int_0^1 b_{\mathbb{K}}(x)K^{\pm}(x_j-x)dx\varepsilon_j \\ =\ & O\left(\frac{\ln^2 n}{n}\right)\sqrt{\int_0^1 K^2(x)dx}+O\left(\frac{\ln^{3/2}n}{\sqrt{n}}\right)\sqrt{\int_0^1 \mathscr{b}_K^2(x)dx}+O\left(\ln^{3/2}n\ n^{-5/4}\right)\sqrt{\int_0^1 K^2(x)dx} \\ & +O\left(\ln^{3/2}n\ n^{-1}\right)\sqrt{\int_0^1 \mathscr{b}_K^2(x)dx}+O\left(\ln^{1/2}n\ n^{-1}\right)\left(\|m\|_2+1\right) \\ =\ & O(n^{-\delta})MISE(K)+O\left(n^{\delta-1}\right) \qquad (57)\end{aligned}$$

whereby the complement of the set of "favorable events" $B^c_{n1,1} \cup B^c_{n3,1} \cup B^c_{n3,0}$ has a probability that decreases like $n^{-\lambda+1/2}$ (Lemmata II.2.2, II.2.4 and inequality (48)).
The results in Sections 2 to 4 yield Propositions **B1** and **B2** of Section 1, thus completing the theorem.

5 VARIATIONS

Since the object of this work is a fixed experimental design, it may be considered natural to examine, instead of MISE, the discrete risk.

$$\widetilde{MISE}(K) = \frac{1}{n}\sum_{i=1}^{n} E\Big[\widetilde{m}_K(x_i) - m(x_i)\Big]^2 \tag{58}$$

The calculations can be performed in an analog way, leaving CV-E[CV] unchanged. In E[CV]-$\widetilde{\text{MISE}}$ the discretization bias drops out, and for $\widetilde{\text{ISE}} - \widetilde{\text{MISE}}$ we obtain $\frac{2}{n} b_{\mathbb{K}}^T \mathbb{K}\varepsilon + \frac{1}{n}\varepsilon^T \mathbb{K}^2 \varepsilon - \frac{1}{n} E[\varepsilon^T \mathbb{K}^2 \varepsilon]$. These polynomial forms have already been handled in Section 2 (CV-E[CV]). It is hence possible to prove the same theorem with the risk defined in (58) in the place of MISE.

Another variation of the statement is obtained when replacing the cross-validation criterion by the sum of squared residuals.

$$RSS(K) = \frac{1}{n}\sum_{i=1}^{n} (\widetilde{m}_K(x_i) - Y_i)^2 \tag{59}$$

This sum is asymptotically biased for MISE,

$$E[RSS(K)] = \frac{1}{n}\sum_{i=1}^{n} b_{\mathbb{K}}^2(x_i) + \sigma^2 \sum_{i,j=1}^{n} K^{\pm 2}(x_i - x_j) - 2\sigma^2 K^{\pm}(0) + \sigma^2$$

The last summand σ^2 is independent of K and cancels out, when subtracting $RSS(K_0) - RSS(K^*)$. So we only have to correct RSS by $2\widetilde{\sigma}^2 K^{\pm}(0)$, where $\widetilde{\sigma}^2$ is an estimator or σ^2. In RSS-E[RSS], all terms which appeared in CV-E[CV] as stemming from the "leave-one-out" method vanish. In contrast, we obtain an additional

$$\left(\widetilde{\sigma}^2 - E\widetilde{\sigma}^2\right) 2K^{\pm}(0) \leq \left(\widetilde{\sigma}^2 - E\widetilde{\sigma}^2\right) \frac{4\sqrt{n}+2}{n}$$

In order to keep the validity of our theorem, $\widetilde{\sigma}^2 - E\widetilde{\sigma}^2$ must satisfy to be $O(n^{-1/2})$ on a set C_n, with $P(C_n^c) = O(n^{-\lambda})$. Even simple non-parametric variance estimators fulfill this condition, for example $\frac{1}{n}\sum_{i=1}^{n}(Y_i - Y_{i+1})^2$ (due to the circular design we can set $Y_{n+1} := Y_1$). Also in E[RSS]-MISE all terms containing the matrix $\mathbb{C}$ vanish. Instead, we have to consider

$$\left(E\widetilde{\sigma}^2 - \sigma^2\right) 2K^{\pm}(0) \leq \left(E\widetilde{\sigma}^2 - \sigma^2\right) \frac{4\sqrt{n}+2}{n}$$

The bias of $\widetilde{\sigma}^2$ should thus not exceed $O(n^{-1/2})$, which is also satisfied by the previously mentioned estimator. ISE-MISE remains unaffected by the exchange of CV with RSS, and we may hence as well prove the theorem for this variation.

Part III
MINIMAX-THEORY

1 SOBOLEV CLASSES and FRACTIONAL DERIVATIVES

In this section we will answer the question whether the minimax properties that are exhibited within Sobolev classes by certain kernel estimators with so called minimax kernels, can also be proven for our density estimation technique which selects the kernel through cross-validation. Sobolev classes are classes of L_2-integrable functions, in our case densities, for which smoothness is measured by the L_2-norm of their β^{th} derivative, $\beta \in \mathbb{N}$. The Sobolev space contains all functions

$$\mathcal{S}_\beta \quad = \quad \left\{ f \in L_2 \Big| \int \left(f^{(\beta)}(x) \right)^2 dx < \infty \right\}. \tag{1}$$

Each Sobolev class $\mathcal{S}_\beta(L)$ is a subset of the Sobolev space $\mathcal{S}_\beta$, on which the L_2-norm of the β-th derivative is bounded by the constant L.

$$\mathcal{S}_\beta(L) \quad = \quad \left\{ f \in L_2 \Big| \int \left(f^{(\beta)}(x) \right)^2 dx \leq L \right\} \tag{2}$$

In Part I it was proven that the proposed method of kernel selection through cross-validation is asymptotically MISE-efficient among all kernel estimators with real, non-negative, unimodal kernel transforms. We will now see that this procedure also attains the asymptotic minimax property among all estimators.

The asymptotic exact minimax risk of density estimators in Sobolev classes was already determined in Efroimovich and Pinsker (1983), Golubev (1991) and (1992), Golubev and Levit (1996) and Schipper (1996). It holds that

$$\begin{aligned} \inf_{\widetilde{f}_n} \sup_{f \in \mathcal{S}_\beta(L)} n^{\frac{2\beta}{2\beta+1}} \; E_f \|\widetilde{f}_n - f\|_2^2 \quad &= \quad \gamma(\beta, L)\Big(1 + o(1)\Big) \\ \text{where} \qquad \gamma(\beta, L) \quad &= \quad (2\beta+1) \left(\frac{\pi(2\beta+1)(\beta+1)}{\beta} \right)^{-\frac{2\beta}{2\beta+1}} L^{\frac{1}{2\beta+1}} \end{aligned}$$

is Pinsker's constant.

At the same time, Efroimovich and Pinsker (1983) proved that kernel density estimators with certain so called minimax kernel functions are efficient in the sense of the minimax risk. Such a minimax kernel function is for instance given by Schipper (1996):

$$K_\beta(x) \quad = \quad \frac{\beta!}{\pi} \sum_{j=1}^{\beta} \frac{\sin^{(j)} x}{(\beta - j)! x^{j+1}} \qquad \text{where} \quad \widehat{K}_\beta(\omega) \quad = \quad \Big(1 - |\omega|^\beta \Big)_+$$

The asymptotic minimax property of kernel CV can be derived in a straight-forward way from the construction of our class of kernel functions, $\mathcal{K} = \{K | \, \widehat{K} \searrow 0\}$, because the bound of Theorem I.2.1 holds uniformly for all f in $\mathcal{S}_\beta(L)$, $\beta > 1/2$. And the class $\mathcal{K}$ is designed so as just to contain the minimax kernel functions K_β for all $\beta \in \mathbb{N}^+ \cup \infty$:

The kernel function selected through CV yields an estimator $\widetilde{f}_{K_0}$, which is asymptotically as good as $\widetilde{f}_{K^*}$, the one that is most adequate for f out of all estimators with $K \in \mathcal{K}$. Which in turn cannot be worse than the estimator with the universal minimax kernel $\widetilde{f}_{K_\beta}$.

However, we are aware of the fact that the characterization of the smoothness of the underlying density function is incomplete when just assigning it to some Sobolev class of L_2-integrable density functions. Recalling the Sobolev criterion,

$$L \;\geq\; \int \left(f^{(\beta)}(x)\right)^2 dx \;=\; \frac{1}{2\pi}\int |\omega^\beta \widehat{f}(\omega)|^2 d\omega,$$

we immediately observe that $\mathcal{S}_\beta(L)$ contains densities which do not lie in $\mathcal{S}_{\beta+1}$, although for suitably chosen $L' < \infty$ and $\varepsilon < 1$, they certainly do satisfy $\frac{1}{2\pi}\int |\omega^{\beta+\varepsilon}\widehat{f}(\omega)|^2 d\omega \leq L'$.
We are hence interested in the question of whether the minimax risk can also be calculated for such generalized Sobolev classes, and if it is even possible to prove the efficiency of the proposed cross-validation technique. For this purpose we will employ the concept of the so called fractional derivative after Riemann and Liouville, thoroughly discussed in Samko (1987):

$$f^{(\beta)}(x) \;=\; \frac{d^\beta}{dx^\beta}f(x) \;=\; \frac{1}{\Gamma(\lceil\beta\rceil-\beta)}\,\frac{d^{\lceil\beta\rceil}}{dx^{\lceil\beta\rceil}}\int t^{\beta-\lceil\beta\rceil}f(x+t)dt$$

with $\lceil x\rceil$ the smallest integer greater than the positive real number x. For $\beta\in\mathbb{N}$, $f^{(\beta)}$ is the β^{th} derivative of f, for $\beta\in\mathbb{R}^+\backslash\mathbb{N}$ it is the β^{th} fractional derivative of f (Samko (1987), p. 137). In case $f^{(\beta)}$ is continuous and L_1-integrable, then $\widehat{f^{(\beta)}}(\omega) = (-i\omega)^\beta\widehat{f}(\omega)$. The other way around, if $(-i\cdot\mathrm{id})^\beta\widehat{f}$ is L_1- or L_2-integrable, the inverse transform from the Fourier into the time domain exists and for our purpose we define:

$$f^{(\beta)}(x) \;:=\; \frac{1}{2\pi}\int(-i\omega)^\beta\widehat{f}(\omega)e^{-ix\omega}dx \tag{3}$$

Existence and uniqueness of the β^{th} fractional derivative of f follow thus from $\frac{1}{2\pi}\int|\omega^\beta\widehat{f}(\omega)|^2d\omega \leq L$, and Parseval's equality gives $\int(f^{(\beta)}(x))^2dx = \frac{1}{2\pi}\int|\omega^\beta\widehat{f}(\omega)|^2d\omega$.
Adopting the idea of Schipper (1996), we find upper and lower bounds for the asymptotic minimax risk in $\mathcal{S}_\beta(L), \beta > 1/2$, which are then shown to converge towards each other. Thereby it will be verified that the minimax risk is determined by $n^{2\beta/(2\beta+1)}\gamma(\beta,L)$, where $\gamma(\beta,L)$ is an exact analogue of Pinsker's constant. Once again, our class $\mathcal{K}$ contains minimax kernels for all Sobolev classes, $\beta > 1/2$, and hence the minimax property of the kernel selection procedure proposed in Part I is proven.
The upper bound results in Section 2 directly from the approximation of the minimax risk in the Fourier domain and is exact, i.e. not asymptotic. Additionally we obtain from this calculation the mentioned minimax kernel functions.
For the lower bound in Section 3, we replace the original problem of estimating a curve by the problem of estimating a finite-dimensional parameter θ (of increasing dimension). A lower bound for the risk of such an estimator may be found by means of the van Trees inequality. In case we were to find a least favorable parametric family $\mathcal{F}_\Theta$ of densities and a least favorable prior distribution Λ on the space of finite-dimensional parameters Θ, such that $f_\theta\in\mathcal{S}_\beta(L)$ with a high probability, the Bayesian risk over $\mathcal{F}_\Theta$ would provide us with a lower bound for the minimax risk on $\mathcal{S}_\beta(L)$.
The result for the lower bound can be seen as a special case of the theorem in Golubev (1992), who yields lower bounds for the quadratic risk of non-parametric estimation problems in a variety of elliptic density classes via Local Asymptotic Normality. Unfortunately the proof

in Golubev (1992) is heavily abbreviated and not easy to retrace. We hope that by our detailed proof, which follows the lines of Schipper (1996), we are able to somehow enlighten the complicated matters.

2 UPPER BOUND

THEOREM III.2.1 Let $\mathcal{S}_\beta(L)$ be the Sobolev class of those L_2-integrable densities, which satisfy $\frac{1}{2\pi}\int |\omega^\beta \widehat{f}(\omega)|^2 d\omega \leq L$ for some constants $\beta > 0$ and $L < \infty$. Then it holds, that

$$\inf_{f_n} \sup_{\mathcal{S}_\beta(L)} n^{\frac{2\beta}{2\beta+1}}\ E_f\|\widetilde{f}_n - f\|_2^2 \quad\leq\quad \gamma(\beta, L)$$

The bound is maintained by a kernel estimator with the minimax kernel K_β, that is the inverse Fourier transform of $\widehat{K}_\beta(\omega) = \left(1 - c_{\min}(L)\cdot|\omega|^\beta\right)_+$, where $c_{\min}(L) = \left(\frac{nL\pi(2\beta+1)(\beta+1)}{\beta}\right)^{-\frac{\beta}{2\beta+1}}$. (For the proof of this theorem, the assumption $\beta > 0$ suffices. Since in contrast, for the lower bound $\beta > 1/2$ is necessary, there arises no gain in generality for minimax statement. But an upper bound for the risk of an estimator is in any case a self-contained result.)

Proof Take $\widetilde{f}_K(x) = \frac{1}{n}\sum K(X_i - x)$ with L_2-integrable kernel function, as before. The steps of approximation will be commented upon during the calculations.

$$\begin{aligned}
& \inf_{\widetilde{f}_n} \sup_{\mathcal{S}_\beta(L)} E_f\|\widetilde{f}_n - f\|_2^2 \\
\leq\ & \inf_{\widetilde{f}_K} \sup_{\mathcal{S}_\beta(L)} E_f\|\widetilde{f}_K - f\|_2^2 \\
=\ & \inf_K \sup_{\mathcal{S}_\beta(L)} \int E_f\left(\widetilde{f}_n(x) - f(x)\right)^2 dx \\
=\ & \inf_K \sup_{\mathcal{S}_\beta(L)} \Big[\ \underbrace{\frac{1}{2\pi}\int |\widehat{f}(\omega)|^2\left(\widehat{K}(\omega) - 1\right)^2 d\omega}_{\substack{\leq\ \sup_\omega \frac{(\widehat{K}(\omega)-1)^2}{|\omega|^{2\beta}}\frac{1}{2\pi}\int|\omega|^{2\beta}|\widehat{f}(\omega)|^2 d\omega \\ \leq\ \sup_\omega \frac{(\widehat{K}(\omega)-1)^2}{|\omega|^{2\beta}}\ L}} + \frac{1}{2\pi n}\int \widehat{K}^2(\omega)\left(1 - |\widehat{f}(\omega)|^2\right) d\omega\Big] \\
\leq\ & \inf_K \Big[\sup_\omega \frac{\left(\widehat{K}(\omega) - 1\right)^2}{|\omega|^{2\beta}}\ L\ + \sup_{\mathcal{S}_\beta(L)} \frac{1}{2\pi n}\int \widehat{K}^2(\omega)\left(1 - |\widehat{f}(\omega)|^2\right) d\omega\Big] \\
=\ & \inf_{c>0} \inf_{|\widehat{K}(\omega)-1|\leq c|\omega|^\beta} \Big[\sup_\omega \frac{\left(\widehat{K}(\omega) - 1\right)^2}{|\omega|^{2\beta}}\ L\ + \sup_{\mathcal{S}_\beta(L)} \frac{1}{2\pi n}\int \widehat{K}^2(\omega)\left(1 - |\widehat{f}(\omega)|^2\right) d\omega\Big] \\
\leq\ & \inf_{c>0}\Big[c^2 L\ + \inf_{|\widehat{K}(\omega)-1|\leq c|\omega|^\beta} \sup_{\mathcal{S}_\beta(L)} \frac{1}{2\pi n}\int \widehat{K}^2(\omega)\left(1 - |\widehat{f}(\omega)|^2\right) d\omega\Big] \\
=\ & \inf_{c>0}\Big[c^2 L\ + \sup_{\mathcal{S}_\beta(L)} \frac{1}{2\pi n}\int \left(1 - c|\omega|^\beta\right)_+^2 \left(1 - |\widehat{f}(\omega)|^2\right) d\omega\Big] \\
\leq\ & \inf_{c>0}\Big[\ \underbrace{c^2 L\ + \frac{1}{2\pi n}\int \left(1 - c|\omega|^\beta\right)_+^2 d\omega}_{\text{uniformly diff.able, strictly konvex}}\ \Big]
\end{aligned}$$

$$
\begin{aligned}
&= c_{\min}^2(L)\cdot L \;+\frac{1}{2\pi n}\int\Big(1-c_{\min}(L)\cdot|\omega|^{\beta}\Big)_+^2\,d\omega\\
&= n^{-\frac{2\beta}{2\beta+1}}\;\gamma(\beta,L)
\end{aligned}
$$

The minimax kernel K_β with $\widehat{K}_\beta(\omega)=\big(1-c_{\min}(L)\cdot|\omega|^\beta\big)_+$, $c_{\min}=\Big(\frac{nL\pi(2\beta+1)(\beta+1)}{\beta}\Big)^{-\frac{\beta}{2\beta+1}}$, is obviously real, non-negative and unimodal, therefore $\in\mathcal{K}$. $\square$

3 LOWER BOUND

We are searching for a lower bound of the minimax risk over the Sobolev class $\mathcal{S}_\beta(L)$. The problem can be reverted to a parametric subset of $\mathcal{S}_\beta$. The question of whether the minimax risk over the subclass coincides with the minimax risk over $\mathcal{S}_\beta(L)$ obviously depends on the difficulty of the estimation problem within the subclass. We achieve our aim using the following construction:
Let us assume $\beta>1/2$ (this assumption is on one hand necessary to the following proof, but on the other hand, $\beta>1/2$ makes the result of Theorem I.2.1 uniform) and let f_0 be the following density from $\mathcal{S}:=\bigcap\limits_{\beta\in\mathbb{R}^+}\mathcal{S}_\beta$ with support $[-1/2,1/2]$:

$$
f_0(x) \;:=\; \begin{cases}\frac{1}{c_a}\exp\Big\{-\frac{a}{(x+1/2)(1/2-x)}\Big\}, & -1/2\le x\le 1/2\\ 0, & \text{otherwise}\end{cases} \tag{4}
$$

with c_a defined so that f_0 is a density, and for technical reason, the constant a satisfying $\int|\widehat{f}_0(\omega)\frac{\sin\omega/2}{\omega/2}|d\omega=2\pi$. Since $\int|\omega^\beta\widehat{f}_0(\omega)|^2d\omega<\infty$ for all $\beta<\infty$, also $\int|\omega^\beta\widehat{f}_0(\omega)|d\omega$ exists for all $\beta<\infty$. Let g_A be the indicator function on $[-A+1/2,A-1/2]$ times the factor $\frac{1}{2A-1}$, i.e.

$$
g_A(x) \;:=\; \frac{1}{2A-1}\;I_{[-A+1/2,A-1/2]}(x). \tag{5}
$$

Then f_0*g_A is a symmetric density within $\mathcal{S}$, that takes the constant value $\frac{1}{2A-1}$ on $[-A+1,A-1]$, and decreases smoothly towards 0 on $[A-1,A]$ and $[-A,-A+1]$. In order to constitute a sufficiently difficult estimation problem departing from this very smooth density, let us add some perturbation functions to f_0*g_A:

$$
\varphi_k(x) \;:=\; \begin{cases}\frac{1}{\sqrt{A}}\cos\frac{k\pi x}{A}\;I_{[-A,A]}(x), & k>0\\ \frac{1}{\sqrt{A}}\sin\frac{k\pi x}{A}\;I_{[-A,A]}(x), & k<0\end{cases} \tag{6}
$$

These perturbations will be weighted by factors θ_k, where $\theta=(\dots,\theta_{-2},\theta_{-1},\theta_1,\theta_2,\dots)$ is (asymptotically) in the set:

$$
\Theta_A(L) \;:=\; \left\{\theta\in\mathbb{R}^\infty\Big|\sum_{k\neq0}|\theta_k|\le A^{-2\beta+1}\text{ and }\sum_{k\neq0}\theta_k^2\left(\frac{k\pi}{A}\right)^{2\beta}\le 4A^2L\right\} \tag{7}
$$

The set $\{f_\theta|\,\theta\in\Theta_A(L)\}$ will from now on be the class of densities under consideration.

$$
f_\theta(x) \;:=\; \frac{1}{b(\theta)}\;f_0*g_A(x)\Big(1+\sum_{k\neq0}\theta_k\varphi_k(x)\Big), \tag{8}
$$

where $b(\theta)$ is the normalization constant. We cannot prove that $\{f_\theta | \theta \in \Theta_A(L)\} \subseteq \mathcal{S}_\beta(L)$, but instead that for all $\varepsilon > 0$ there exists an $A_\varepsilon < \infty$, so that for every $A \geq A_\varepsilon$ the following holds: $\sup_{\theta\in\Theta_A(L)} \|f_\theta^{(\beta)}\|_2^2 \leq L + \varepsilon$. If we want to let $\varepsilon \longrightarrow 0$, we have to let $A \longrightarrow \infty$, which will be seen in the proof of Theorem III.3.1.

THEOREM III.3.1 Let f_θ and $\Theta_A(L)$ be defined as above. Then, as $A \longrightarrow \infty$:

$$\sup_{\theta\in\Theta_A(L)} \|f_\theta^{(\beta)}\|_2^2 \quad \leq \quad L + o(1)$$

Since the proof of this theorem is purely technical and unpleasantly lengthy, we proceed directly with the derivation of the lower bound for the minimax risk. Nevertheless we want to give the proof in the context of this work (to be found in Subsection 3.1) because until now it is not yet available in the literature.
The next step leading to the lower bound requires the definition of a prior distribution Λ, which is done accordingly to Schipper (1996), so as to yield a parameter θ of finite dimension: Let $\varepsilon > 0$, $W > 0$ and $\sigma_k^2 > 0$

$$\lambda(\theta) \quad = \quad \prod_{0<|k|<W} \lambda_k(\theta_k) \prod_{|k|\geq W} \delta_0(\theta_k), \tag{9}$$

where $\delta_0(.)$ is the Dirac function on 0, and for $|k| < W$: $\lambda_k(\theta_k)$ are absolutely continuous densities with $E\theta_k^2 = \sigma_k^2$, $\theta_k^2 \leq G^2\sigma_k^2 \quad (\Lambda{-}f.s.)$, and the Fisher information $I_k := \int \frac{\lambda_k'^2(\theta_k)}{\lambda_k(\theta_k)}\, d\theta_k$ $\leq (1+\varepsilon)\sigma_k^{-2}$ (with respect to the translation group $\{\lambda_k(.-u)|u \in \mathbb{R}\}$). (These conditions are satisfied, for example, by independent bounded, zero mean random variables $\sigma_k\xi_k$, $|k| < W$, with $|\xi_k| < G$, $E\xi_k^2 = 1$ and the Fisher-information of the density of ξ_k smaller than $1+\varepsilon$.)
Let us set

$$\begin{aligned} W \quad &= \quad \frac{A}{\pi}\Big(\frac{L(1-\varepsilon)n(2\beta+1)(\beta+1)\pi}{\beta}\Big)^{\frac{1}{2\beta+1}} \\ \sigma_k^2 \quad &= \quad \frac{4A}{n}\Big(\Big|\frac{W}{k}\Big|^\beta - 1\Big)_+ \end{aligned} \tag{10}$$

As W grows with $n \longrightarrow \infty$, the dimension of the parameter θ will tend to infinity, allowing for more and more perturbation functions φ_k in the definition of f_θ. At the end of Section 3 it will be shown that σ_k^2 and W of this form approximately maximize the lower bound of the minimax risk for the prior distribution Λ.
Since Λ is not supported on $\Theta_A(L)$, we will have to show that at least the probability of $\theta \in \Theta_A(L)$ grows with $n \longrightarrow \infty$:
First consider that λ has a bounded support, $|\theta_k| \leq G\sigma_k$ for $|k| < W$, and else $\theta_k = 0$. With the above construction of σ_k^2 and W, letting $A \sim \ln n$, condition $\sum |\theta_k| \leq A^{-2\beta+1}$ is fulfilled for n sufficiently large. Lemma III.3.2 takes care of $\sum \theta_k^2 \left(\frac{k\pi}{A}\right)^{2\beta} \leq 4A^2L$.

Lemma III.3.2 For the prior distribution Λ defined above, with W and σ_k^2 as in (10), it holds that for $n \longrightarrow \infty$:

$$P_\lambda\Big(\theta \notin \Theta_A(L)\Big) \quad = \quad o(n^{-1})$$

Proof Since W converges towards ∞ when $n \longrightarrow \infty$, we have:

$$\begin{aligned}
E_\lambda \sum_{k\neq 0} \theta_k^2 \left(\frac{k\pi}{A}\right)^{2\beta} &= \sum_{0<|k|<W} \sigma_k^2 \left(\frac{k\pi}{A}\right)^{2\beta} \\
&= \frac{4\pi^{2\beta}}{A^{2\beta-1}n} \sum_{0<k<W} k^{2\beta} \left(\left(\frac{W}{k}\right)^{\beta} - 1\right) \\
&= \frac{4\pi^{2\beta}W^{2\beta+1}}{A^{2\beta-1}n} \cdot \frac{\beta}{(2\beta+1)(\beta+1)} \Big(1+o(1)\Big) \\
&= 4A^2L(1-\varepsilon)\Big(1+o(1)\Big).
\end{aligned}$$

Hence, for sufficiently large n, $E_\lambda \sum \theta_k^2 \left(\frac{k\pi}{A}\right)^{2\beta} \leq 4A^2L(1-\varepsilon/2)$. The result follows from the Höffding-inequality (Pollard (1984), p. 191).

$$\begin{aligned}
P_\lambda\Big(\theta \notin \Theta_A(L)\Big) &= P_\Lambda \left(\sum_{k\neq 0} \theta_k^2 \left(\frac{k\pi}{A}\right)^{2\beta} > 4A^2L\right) \\
&\leq P_\Lambda \left(\sum_{k\neq 0} \left(\theta_k^2 - E_\Lambda \theta_k^2\right) \left(\frac{k\pi}{A}\right)^{2\beta} > \frac{\varepsilon}{2}\, 4A^2L\right) \\
&\leq \exp\left\{-\frac{(4A^2L)^2\varepsilon^2}{2Q}\right\}
\end{aligned}$$

where

$$\begin{aligned}
Q &= \sum_{0<|k|<W} \left(\left(\frac{k\pi}{A}\right)^{2\beta} 2(G^2+1)\sigma_k^2\right)^2 \\
&= 4(G^2+1)^2 \sum_{0<|k|<W} \left(\frac{k\pi}{A}\right)^{4\beta} \sigma_k^4 \\
&= 4(G^2+1)^2\ 2\sum_{0<k<W} \left(\frac{k\pi}{A}\right)^{4\beta} \frac{4^2A^2}{n^2} \left(\left(\frac{W}{k}\right) - 1\right)^2 \\
&\leq 8(G^2+1)^2\ \frac{4^2A^2}{n^2} \left(\frac{\pi}{A}\right)^{4\beta} W^{4\beta+1}\Big(1+o(1)\Big) \\
&\leq 128(G^2+1)^2\ A^3 n^{-\frac{1}{2\beta+1}} \pi^{\frac{2\beta}{2\beta+1}} \left(\frac{L(1-\varepsilon)(2\beta+1)(\beta+1)}{\beta}\right)^{\frac{4\beta+1}{2\beta+1}} \Big(1+o(1)\Big)
\end{aligned}$$

Thus we obtain:

$$P_\lambda\Big(\theta \notin \Theta_A(L)\Big) \leq \exp\left\{-\text{const}\ A\,\varepsilon^2 n^{\frac{1}{2\beta+1}}\right\} \qquad \square$$

THEOREM III.3.3 For $L < \infty$, $\beta > 1/2$ and $\gamma(\beta, L)$ equal to Pinsker's constant we have:

$$\liminf_{n\to\infty} \inf_{\widetilde{f}_n} \sup_{f\in\mathcal{S}_\beta(L)} n^{\frac{2\beta}{2\beta+1}}\ E_f\|\widetilde{f}_n - f\|_2^2 \geq \gamma(\beta, L)$$

Proof Let us at first reduce the supremum of the risk by restricting the set of density functions. According to Theorem III.3.1 we know that for $A \sim \ln n$, $\lim_{A\to\infty}\{f_\theta | \theta \in \Theta_A(L)\} \subseteq \mathcal{S}_\beta(L)$.

$$\liminf_{n\to\infty} \inf_{\widetilde{f}_n} \sup_{f\in\mathcal{S}_\beta(L)} E_f\|\widetilde{f}_n - f\|_2^2 \;\geq\; \liminf_{n\to\infty} \inf_{\widetilde{f}_n} \sup_{\theta\in\Theta_A(L)} E_{f_\theta}\|\widetilde{f}_n - f_\theta\|_2^2 \tag{11}$$

For any fixed A, we find a lower bound for the supremum over $\Theta_A(L)$ through the Bayesian risk with respect to Λ.

$$\inf_{\widetilde{f}_n} \sup_{\theta\in\Theta_A(L)} E_{f_\theta}\|\widetilde{f}_n - f_\theta\|_2^2 \;\geq\; \inf_{\widetilde{f}_n} \int_{\Theta_A(L)} E_{f_\theta}\|\widetilde{f}_n - f_\theta\|_2^2 \, d\Lambda(\theta) \tag{12}$$

In inequality (19) of the proof of Theorem III.3.1 it will be shown that $1 - A^{-3/2} \leq b(\theta) \leq 1 + A^{-3/2}$. (20) and (21) are also feasible for $\beta = 0$, leading to $\|f_0 * g_A\|_2 \leq (2A-1)^{-1}\|f_0\|_2$. Furthermore, because of orthonormality $\|\sum \theta_k\varphi_k\|_2^2 = \sum\theta_k^2 \leq \sum|\theta_k| \leq A^{-2\beta+1}$, (32). So we can derive, for all $\theta \in \Theta_A(L)$:

$$\begin{aligned}
\|f_\theta\|_2 &= \frac{1}{b(\theta)}\Big\|f_0 * g_A\Big(1 - \sum_{k\neq 0}\theta_k\varphi_k\Big)\Big\|_2 \\
&\leq \frac{1}{b(\theta)}\Big\|f_0 * g_A\Big\|_2 + \frac{1}{b(\theta)}\Big\|f_0 * g_A\sum_{k\neq 0}\theta_k\varphi_k\Big\|_2 \\
&\leq \frac{\text{const.}}{A} \;=:\; \frac{1}{A_0}
\end{aligned}$$

Because the set of all densities with $\|f\|_2 \leq 1/A_0$ is convex, we may in (13) also restrict the set estimators to $\|\widetilde{f}_n\|_2 \leq 1/A_0$ without increasing the supremum.

$$\begin{aligned}
\inf_{\widetilde{f}_n}\int_{\Theta_A(L)} E_{f_\theta}\|\widetilde{f}_n - f_\theta\|_2^2\, d\Lambda(\theta) &= \inf_{\|\widetilde{f}_n\|_2^2\leq A_0^{-2}}\int_{\Theta_A(L)} E_{f_\theta}\|\widetilde{f}_n - f_\theta\|_2^2\, d\Lambda(\theta) && (13)\\
&\geq \inf_{\|\widetilde{f}_n\|_2^2\leq A_0^{-2}}\int E_{f_\theta}\|\widetilde{f}_n - f_\theta\|_2^2\, d\Lambda(\theta) - \frac{4}{A_0^2}P_\Lambda\Big(\theta \notin \Theta_A(L)\Big) && (14)\\
&\geq \inf_{\|\widetilde{f}_n\|_2^2\leq A_0^{-2}}\int E_{f_\theta}\|\widetilde{f}_n - f_\theta\|_2^2\, d\Lambda(\theta) \;+\; o(n^{-1}) \\
&\geq \inf_{\widetilde{f}_n} E_\lambda E_{f_\theta}\|\widetilde{f}_n - f_\theta\|_2^2 \;+\; o(n^{-1}) && (15)
\end{aligned}$$

Due to $\|f_\theta\|_2^2 \leq A_0^{-2}$ and $\|\widetilde{f}_n\|_2^2 \leq A_0^{-2}$ it holds in (14): $\|\widetilde{f}_n - f_\theta\|_2^2 \leq 4A_0^{-2}$. In (15) we return to the complete class of estimators.

Since f_θ has bounded support, i.e. $[-A, A]$, it is equivalent, as regards the quadratic risk, either to estimate the function f_θ in the time domain or its Fourier coefficients. ($\widehat{f_\theta}(0) = 1$ is known)

$$\begin{aligned}
& E_\lambda E_{f_\theta}\|\widetilde{f}_n - f_\theta\|_2^2 \\
=\; & E_\lambda E_{f_\theta}\frac{1}{2A}\sum_{\kappa\neq 0}\Big|\widehat{\widetilde{f}_n}\Big(\frac{\kappa\pi}{A}\Big) - \widehat{f}_\theta\Big(\frac{\kappa\pi}{A}\Big)\Big|^2 \\
=\; & E_\lambda E_{f_\theta}\frac{1}{2A}\sum_{\kappa\neq 0}\text{Re}^2\Big(\widehat{\widetilde{f}_n}\Big(\frac{\kappa\pi}{A}\Big) - \widehat{f}_\theta\Big(\frac{\kappa\pi}{A}\Big)\Big) + \text{Im}^2\Big(\widehat{\widetilde{f}_n}\Big(\frac{\kappa\pi}{A}\Big) - \widehat{f}_\theta\Big(\frac{\kappa\pi}{A}\Big)\Big) \\
=\; & E_\lambda E_{f_\theta}\frac{1}{2A}\sum_{\kappa\neq 0}\Big(\text{Re}\widehat{\widetilde{f}_n}\Big(\frac{\kappa\pi}{A}\Big) - \text{Re}\widehat{f}_\theta\Big(\frac{\kappa\pi}{A}\Big)\Big)^2 + \Big(\text{Im}\widehat{\widetilde{f}_n}\Big(\frac{\kappa\pi}{A}\Big) - \text{Im}\widehat{f}_\theta\Big(\frac{\kappa\pi}{A}\Big)\Big)^2 \qquad (16)
\end{aligned}$$

The van Trees inequality (Gill und Levit (1995), see also Appendix A.4) may now be applied on every single summand. For technical reason, the real parts are derived with respect to $\theta_{|\kappa|}$, while the imaginary ones are derived with respect to $\theta_{-|\kappa|}$.

$$
\begin{aligned}
E_\lambda E_{f_\theta}\left[\mathrm{Re}\widetilde{\widehat{f}_n}\left(\frac{\kappa\pi}{A}\right)-\mathrm{Re}\widehat{f}_\theta\left(\frac{\kappa\pi}{A}\right)\right]^2 &\geq \frac{E_\lambda^2\left[\partial\,\mathrm{Re}\widehat{f}_\theta\left(\frac{\kappa\pi}{A}\right)/\partial\theta_{|\kappa|}\right]}{nE_\lambda I_{f_\theta}(\theta_{|\kappa|})+I_{|\kappa|}} \\
E_\lambda E_{f_\theta}\left[\mathrm{Im}\widetilde{\widehat{f}_n}\left(\frac{\kappa\pi}{A}\right)-\mathrm{Im}\widehat{f}_\theta\left(\frac{\kappa\pi}{A}\right)\right]^2 &\geq \frac{E_\lambda^2\left[\partial\,\mathrm{Im}\widehat{f}_\theta\left(\frac{\kappa\pi}{A}\right)/\partial\theta_{-|\kappa|}\right]}{nE_\lambda I_{f_\theta}(\theta_{-|\kappa|})+I_{-|\kappa|}}
\end{aligned}
$$

where we denote $I_{f_\theta}(\theta_\kappa)=\int\frac{(\partial f_\theta(x)/\partial\theta_\kappa)^2}{f_\theta(x)}\,dx$. I_κ is the "Fisher information" of λ_κ and by construction $\leq(1+\varepsilon)\sigma_\kappa^{-2}$ for $|\kappa|<W$ and $=\infty$ for $|\kappa|\geq W$, respectively. Hence all summands with $|\kappa|\geq W$ vanish from the sum. Approximations for $I_{f_\theta}(\theta_\kappa)$, $\partial\,\mathrm{Re}\,\widehat{f}_\theta\left(\frac{\kappa\pi}{A}\right)/\partial\theta_{|\kappa|}$ und $\partial\,\mathrm{Im}\,\widehat{f}_\theta\left(\frac{\kappa\pi}{A}\right)/\partial\theta_{-|\kappa|}$ are available from

Lemma III.3.4 For $A\longrightarrow\infty$:

$$
I_{f_\theta}(\theta_\kappa)=\frac{1+o(1)}{2A}\qquad\frac{\partial\,\mathrm{Re}\widehat{f}_\theta\left(\frac{\kappa\pi}{A}\right)}{\partial\theta_{|\kappa|}}=\frac{1+o(1)}{2\sqrt{A}}\qquad\frac{\partial\,\mathrm{Im}\widehat{f}_\theta\left(\frac{\kappa\pi}{A}\right)}{\partial\theta_{-|\kappa|}}=\frac{1+o(1)}{2\sqrt{A}}
$$

with $o(1)$ independent of κ and θ_κ. The proof is postponed to the Subsection 3.1. From (15) completed by (16), the van Trees approximation and Lemma III.3.4 we thus have:

$$
\begin{aligned}
\inf_{\widetilde{f}_n}E_\lambda E_{f_\theta}\|\widetilde{f}_n-f_\theta\|_2^2 &= \inf_{\widetilde{f}_n}E_\lambda E_{f_\theta}\frac{1}{2A}\sum_{\kappa\neq0}\left|\widetilde{\widehat{f}_n}\left(\frac{\kappa\pi}{A}\right)-\widehat{f}_\theta\left(\frac{\kappa\pi}{A}\right)\right|^2 \\
&= E_\lambda E_{f_\theta}\frac{1}{2A}\sum_{\kappa\neq0}\left(\mathrm{Re}\widetilde{\widehat{f}_n}\left(\frac{\kappa\pi}{A}\right)-\mathrm{Re}\widehat{f}_\theta\left(\frac{\kappa\pi}{A}\right)\right)^2+\left(\mathrm{Im}\widetilde{\widehat{f}_n}\left(\frac{\kappa\pi}{A}\right)-\mathrm{Im}\widehat{f}_\theta\left(\frac{\kappa\pi}{A}\right)\right)^2 \\
&\geq \frac{1}{2A}\sum_{\kappa\neq0}\frac{E_\lambda^2\left[\partial\,\mathrm{Re}\widehat{f}_\theta\left(\frac{\kappa\pi}{A}\right)/\partial\theta_{|\kappa|}\right]}{nE_\lambda I_{f_\theta}(\theta_{|\kappa|})+I_{|\kappa|}}+\frac{E_\lambda^2\left[\partial\,\mathrm{Im}\widehat{f}_\theta\left(\frac{\kappa\pi}{A}\right)/\partial\theta_{-|\kappa|}\right]}{nE_\lambda I_{f_\theta}(\theta_{-|\kappa|})+I_{-|\kappa|}} \\
&= \frac{1}{2A}\sum_{0<|\kappa|<W}\frac{\left(\frac{1+o(1)}{2\sqrt{A}}\right)^2}{n\,\frac{1+o(1)}{2A}+(1+\varepsilon)\sigma_{|\kappa|}^{-2}}+\frac{\left(\frac{1+o(1)}{2\sqrt{A}}\right)^2}{n\,\frac{1+o(1)}{2A}+(1+\varepsilon)\sigma_{-|\kappa|}^{-2}} \\
&= \frac{1+o(1)}{2A(1+\varepsilon)}\sum_{0<|\kappa|<W}\frac{1}{n+2A\sigma_\kappa^{-2}} \qquad (17)
\end{aligned}
$$

All sums obtained from W and σ_κ^2 through (17), i.e. from a prior distribution Λ satisfying hypotheses (10) and Lemma III.3.2, are thus lower bounds of the minimax risk.
What we are searching for is a bound as large as possible, we hence maximize (17) subject to the constraint $\sum\sigma_\kappa^2\left(\frac{\kappa\pi}{A}\right)^{2\beta}\leq(1-\varepsilon)4A^2L$, such that $P(\theta\notin\Theta_A(L))=o(n^{-1})$ remains valid. The solution to this problem is W and σ_κ^2 from (10). The maximum in (17) can be approximated as follows:

$$
\frac{1}{2A(1+\varepsilon)n}\sum_{0<|\kappa|<W}\left(\left|\frac{W}{\kappa}\right|^\beta-1\right)\left|\frac{\kappa}{W}\right|^\beta
$$

$$
\begin{aligned}
&= \frac{1}{A(1+\varepsilon)n} \sum_{0<\kappa<W} \left(1-\left(\frac{\kappa}{W}\right)^{\beta}\right) \\
&= \frac{1}{A(1+\varepsilon)n} \frac{\beta}{\beta+1} W\Big(1+o(1)\Big) \\
&= (2\beta+1)\left(\frac{(2\beta+1)(\beta+1)\pi}{\beta n}\right)^{-\frac{2\beta}{2\beta+1}} L^{\frac{1}{2\beta+1}} \frac{(1-\varepsilon)^{\frac{1}{2\beta+1}}}{1+\varepsilon}\Big(1+o(1)\Big) \\
&= n^{-\frac{2\beta}{2\beta+1}} \gamma(\beta,L) \frac{(1-\varepsilon)^{\frac{1}{2\beta+1}}}{1+\varepsilon}\Big(1+o(1)\Big) \qquad (18)
\end{aligned}
$$

Combining (11) with (12), (15), (17) and (18), we obtain the required result:

$$
\liminf_{n\to\infty} \inf_{\widetilde{f}_n} \sup_{f\in\mathcal{S}_\beta(L)} n^{\frac{2\beta}{2\beta+1}} E_f\|\widetilde{f}_n - f\|_2^2 \geq \gamma(\beta,L) \qquad \square
$$

3.1 Postponed Proofs

Proof of **Theorem III.3.1** For f_θ defined in equation (8), it holds that

$$
\begin{aligned}
\|f_\theta^{(\beta)}\|_2 &= \frac{1}{b(\theta)}\Big\|(f_0 * g_A)^{(\beta)} + \Big(f_0 * g_A \sum_{k\neq 0}\theta_k\varphi_k\Big)^{(\beta)}\Big\|_2 \\
&\leq \frac{1}{b(\theta)}\Big\|(f_0 * g_A)^{(\beta)}\Big\|_2 + \frac{1}{b(\theta)}\Big\|\Big(f_0 * g_A \sum_{k\neq 0}\theta_k\varphi_k\Big)^{(\beta)}\Big\|_2
\end{aligned}
$$

$b(\theta)$, $\|(f_0 * g_A)^{(\beta)}\|_2^2$ and $\|(f_0 * g_A \sum\theta_k\varphi_k)^{(\beta)}\|_2^2$ are then considered one by one. Remember definition (6): $\varphi_k(x) = A^{-1/2}\cos(\pi k/A)I(|x|\leq A)$ for $k>0$, and the same with sine for $k<0$. Take first the normalizing constant $b(\theta)$:

$$
\begin{aligned}
b(\theta) &= \int_{-A}^{A} f_0 * g_A(x)\left(1+\sum_{k\neq 0}\theta_k\varphi_k(x)\right) dx \\
&= 1+\sum_{k\neq 0}\theta_k \int_{-A}^{A} f_0 * g_A(x)\varphi_k(x)dx \\
&= 1+\sum_{k>0}\theta_k \int_{-A}^{A} f_0 * g_A(x)\frac{1}{\sqrt{A}}\cos\frac{k\pi x}{A}\,dx + \sum_{k<0}\theta_k \int_{-A}^{A} f_0 * g_A(x)\frac{1}{\sqrt{A}}\sin\frac{k\pi x}{A}\,dx \\
&= 1+\sum_{k>0}\theta_k \int_{-A}^{A} f_0 * g_A(x)\frac{1}{\sqrt{A}}\cos\frac{k\pi x}{A}\,dx + 0 \\
&= 1+\frac{1}{\sqrt{A}}\sum_{k>0}\theta_k\left[\int_{-A}^{A}\frac{1}{2A-1}\cos\frac{k\pi x}{A}\,dx - 2\int_{A-1}^{A}\left(\frac{1}{2A-1} - f_0 * g_A(x)\right)\cos\frac{k\pi x}{A}\,dx\right] \\
&= 1+\frac{1}{\sqrt{A}}\sum_{k>0}\theta_k\left[0 - 2\int_{A-1}^{A}\left(\frac{1}{2A-1} - f_0 * g_A(x)\right)\cos\frac{k\pi x}{A}\,dx\right]
\end{aligned}
$$

For the second term on the right-hand side we have:

$$
\frac{2}{\sqrt{A}}\Big|\sum_{k>0}\theta_k\int_{A-1}^{A}\left(\frac{1}{2A-1} - f_0 * g_A(x)\right)\cos\frac{k\pi x}{A}\,dx\Big| \leq \frac{2}{\sqrt{A}}\sum_{k>0}|\theta_k| \int_{A-1}^{A}\frac{1}{2A-1}\Big|\cos\frac{k\pi x}{A}\Big|\,dx
$$

$$\leq \frac{2}{\sqrt{A}(2A-1)} \sum_{k>0} |\theta_k|$$
$$\leq \frac{1}{\sqrt{A}(A-1/2)} A^{-2\beta+1},$$

so that for $\beta > 1/2$ and A sufficiently large, it follows that

$$1 - A^{-3/2} \leq b(\theta) \leq 1 + A^{-3/2} \tag{19}$$

f_0 and g_A have, in addition to L_2-integrability, a bounded support. So instead of the L_2-norm of $f_0 * g_A$ in the time domain, by Parseval's equality we may as well study the ℓ_2-norm of their Fourier coefficients. For any function $f : [-A, A] \longrightarrow \mathbb{R}$ that can be expanded into a Fourier series (see Appendix A1 for sufficient conditions and calculation rules), we define

$$\begin{aligned}
\widehat{f}\Big(\frac{\kappa\pi}{A}\Big) &:= \int_{-A}^{A} f(x) e^{ix\frac{\kappa\pi}{A}}\, dx \qquad \text{for } \kappa \in \mathbb{Z} \\
\|\widehat{f}\|_{\ell_2}^2 &:= \frac{1}{2A} \sum_{\kappa\in\mathbb{Z}} \Big|\widehat{f}\Big(\frac{\kappa\pi}{A}\Big)\Big|^2 \qquad \text{such that} \qquad \|f\|_2^2 = \|\widehat{f}\|_{\ell_2}^2 \\
\text{and} \quad \|\widehat{f}\|_{\ell_1} &:= \frac{1}{2A} \sum_{\kappa\in\mathbb{Z}} \Big|\widehat{f}\Big(\frac{\kappa\pi}{A}\Big)\Big|
\end{aligned}$$

Then we can approximate

$$\begin{aligned}
\Big\|(f_0 * g_A)^{(\beta)}\Big\|_2^2 &= \int_{-A}^{A} \Big|(f_0 * g_A)^{(\beta)}(x)\Big|^2 dx \\
&= \frac{1}{2A} \sum_{\kappa\in\mathbb{Z}} \Big|\Big(\frac{\kappa\pi}{A}\Big)^\beta \widehat{f_0 * g_A}\Big(\frac{\kappa\pi}{A}\Big)\Big|^2 \\
&= \frac{1}{2A} \sum_{\kappa\in\mathbb{Z}} \Big|\Big(\frac{\kappa\pi}{A}\Big)^\beta \widehat{f}_0\Big(\frac{\kappa\pi}{A}\Big)\widehat{g}_A\Big(\frac{\kappa\pi}{A}\Big)\Big|^2 \\
&= \frac{1}{2A} \sum_{\kappa\in\mathbb{Z}} \Big|\Big(\frac{\kappa\pi}{A}\Big)^\beta \widehat{f}_0\Big(\frac{\kappa\pi}{A}\Big) \frac{2\sin\frac{(A-1/2)\kappa\pi}{A}}{(2A-1)\frac{\kappa\pi}{A}}\Big|^2 \\
&= \frac{1}{2A} \sum_{\kappa\in\mathbb{Z}} \Big|\Big(\frac{\kappa\pi}{A}\Big)^\beta \widehat{f}_0\Big(\frac{\kappa\pi}{A}\Big) \frac{\sin\big(\kappa\pi - \frac{\kappa\pi}{2A}\big)}{(2A-1)\frac{\kappa\pi}{2A}}\Big|^2 \\
&= \frac{1}{2A} \sum_{\kappa\in\mathbb{Z}} \Big|\Big(\frac{\kappa\pi}{A}\Big)^\beta \widehat{f}_0\Big(\frac{\kappa\pi}{A}\Big) \frac{(-1)^{\kappa+1}\sin\frac{\kappa\pi}{2A}}{(2A-1)\frac{\kappa\pi}{2A}}\Big|^2
\end{aligned} \tag{20}$$

The function $x^{-1}\sin x$ is in absolute value smaller than 1, so we can continue at inequality number (20):

$$\begin{aligned}
\Big\|(f_0 * g_A)^{(\beta)}\Big\|_2^2 &\leq \Big(\frac{1}{2A-1}\Big)^2 \frac{1}{2A} \sum_{\kappa\in\mathbb{Z}} \Big|\Big(\frac{\kappa\pi}{A}\Big)^\beta \widehat{f}_0\Big(\frac{\kappa\pi}{A}\Big)\Big|^2 \\
&= \Big(\frac{1}{2A-1}\Big)^2 \|f_0^{(\beta)}\|_2^2
\end{aligned} \tag{21}$$

The consideration of the last and most important term $\|(f_0 * g_A \cdot \sum \theta_k \varphi_k)^{(\beta)}\|_2^2$ is a bit more tricky.

$$
\begin{aligned}
&\Big\|\Big(f_0 * g_A \cdot \sum_{k\neq 0} \theta_k \varphi_k\Big)^{(\beta)}\Big\|_2^2 \\
= \;& \frac{1}{2A}\sum_{\kappa\in\mathbb{Z}}\Big|\Big(\frac{\kappa\pi}{A}\Big)^{\beta}\left(\frac{1}{2A}\;\widehat{f_0 * g_A} * \sum_{k\neq 0}\theta_k\widehat{\varphi}_k\right)\Big(\frac{\kappa\pi}{A}\Big)\Big|^2 \\
= \;& \frac{1}{2A}\sum_{\kappa\in\mathbb{Z}}\Big|\Big(\frac{\kappa\pi}{A}\Big)^{\beta}\frac{1}{2A}\sum_{\lambda\in\mathbb{Z}}\widehat{f_0 * g_A}\Big(\frac{\lambda\pi}{A}\Big)\sum_{k\neq 0}\theta_k\widehat{\varphi}_k\Big(\frac{(\kappa-\lambda)\pi}{A}\Big)\Big|^2 \\
= \;& \frac{1}{2A}\sum_{\kappa\in\mathbb{Z}}\Big|\Big(\frac{\kappa\pi}{A}\Big)^{\beta}\frac{1}{2A}\sum_{\lambda\in\mathbb{Z}}\widehat{f_0 * g_A}\Big(\frac{\lambda\pi}{A}\Big)\sum_{k\neq 0}\theta_k\widehat{\varphi}_k\Big(\frac{(\kappa-\lambda)\pi}{A}\Big)\Big| \\
& \quad\times\Big|\Big(\frac{\kappa\pi}{A}\Big)^{\beta}\frac{1}{2A}\sum_{\mu\in\mathbb{Z}}\widehat{f_0 * g_A}\Big(\frac{\mu\pi}{A}\Big)\sum_{k\neq 0}\theta_k\widehat{\varphi}_k\Big(\frac{(\kappa-\mu)\pi}{A}\Big)\Big| \\
\leq \;& \frac{1}{2A}\sum_{\lambda\in\mathbb{Z}}\Big|\widehat{f_0 * g_A}\Big(\frac{\lambda\pi}{A}\Big)\Big|\;\frac{1}{2A}\sum_{\mu\in\mathbb{Z}}\Big|\widehat{f_0 * g_A}\Big(\frac{\mu\pi}{A}\Big)\Big|\;\frac{1}{2A}\sum_{\kappa\in\mathbb{Z}}\left[\Big|\Big(\frac{\kappa\pi}{A}\Big)^{\beta}\sum_{k\neq 0}\theta_k\widehat{\varphi}_k\Big(\frac{(\kappa-\lambda)\pi}{A}\Big)\Big|\right. \\
& \quad\left.\times\Big|\Big(\frac{\kappa\pi}{A}\Big)^{\beta}\sum_{k\neq 0}\theta_k\widehat{\varphi}_k\Big(\frac{(\kappa-\mu)\pi}{A}\Big)\Big|\right] \\
\leq \;& \frac{1}{2A}\sum_{\lambda\in\mathbb{Z}}\Big|\widehat{f_0 * g_A}\Big(\frac{\lambda\pi}{A}\Big)\Big|\;\frac{1}{2A}\sum_{\mu\in\mathbb{Z}}\Big|\widehat{f_0 * g_A}\Big(\frac{\mu\pi}{A}\Big)\Big|\;\sqrt{\frac{1}{2A}\sum_{\kappa\in\mathbb{Z}}\Big|\Big(\frac{\kappa\pi}{A}\Big)^{\beta}\sum_{k\neq 0}\theta_k\widehat{\varphi}_k\Big(\frac{(\kappa-\lambda)\pi}{A}\Big)\Big|^2} \\
& \quad\times\sqrt{\frac{1}{2A}\sum_{\kappa\in\mathbb{Z}}\Big|\Big(\frac{\kappa\pi}{A}\Big)^{\beta}\sum_{k\neq 0}\theta_k\widehat{\varphi}_k\Big(\frac{(\kappa-\mu)\pi}{A}\Big)\Big|^2} \qquad (22)
\end{aligned}
$$

At this point we apply Lemma III.3.5, which states that:

$$
\begin{aligned}
\frac{1}{2A}\sum_{\kappa\in\mathbb{Z}}\Big|\Big(\frac{\kappa\pi}{A}\Big)^{\beta}\sum_{k\neq 0}\theta_k\widehat{\varphi}_k\Big(\frac{(\kappa-\lambda)\pi}{A}\Big)\Big|^2 \;&=\; 4LA^2+(1+4L)o(A^2)\sum_{i=1}^{2\lfloor\beta\rfloor+1}c'_{(i;\beta)}\Big|\frac{\lambda\pi}{A}\Big|^{i} \\
&\quad + c'_{(2\lceil\beta\rceil;\beta)}\Big|\frac{\lambda\pi}{A}\Big|^{2\lceil\beta\rceil}o(A^2)
\end{aligned}
$$

where $c'_{(i;\beta)}$ and $c'_{(2\lceil\beta\rceil;\beta)}$ are constants defined in Lemma III.3.5, that are independent from θ and A, and $o(A^2)$ is of order A for $\beta \in \mathbb{N}$ and of order $A^{1+\lceil\beta\rceil-\beta}$ for $\beta \in \mathbb{R}^+\backslash\mathbb{N}$, respectively. Hence continuing at inequality number (22):

$$
\begin{aligned}
&\Big\|\Big(f_0 * g_A \cdot \sum_{k\neq 0} \theta_k \varphi_k\Big)^{(\beta)}\Big\|_2^2 \\
\leq \;& \frac{1}{2A}\sum_{\lambda\in\mathbb{Z}}\Big|\widehat{f_0 * g_A}\Big(\frac{\lambda\pi}{A}\Big)\Big|\;\frac{1}{2A}\sum_{\mu\in\mathbb{Z}}\Big|\widehat{f_0 * g_A}\Big(\frac{\mu\pi}{A}\Big)\Big| \\
& \quad\times\sqrt{4LA^2+(1+4L)o(A^2)\sum_{i=1}^{2\lfloor\beta\rfloor+1}c'_{(i;\beta)}\Big|\frac{\lambda\pi}{A}\Big|^{i}+c'_{(2\lceil\beta\rceil;\beta)}\Big|\frac{\lambda\pi}{A}\Big|^{2\lceil\beta\rceil}o(A^2)}
\end{aligned}
$$

$$
\begin{aligned}
&\times \sqrt{4LA^2 + (1+4L)o(A^2) \sum_{i=1}^{2\lfloor\beta\rfloor+1} c'_{(i;\beta)} \Big|\frac{\mu\pi}{A}\Big|^i + c'_{(2\lceil\beta\rceil;\beta)} \Big|\frac{\mu\pi}{A}\Big|^{2\lceil\beta\rceil} o(A^2)} \\
= &\left\{ \frac{1}{2A} \sum_{\lambda\in\mathbb{Z}} \Big| \widehat{f_0 * g_A}\Big(\frac{\lambda\pi}{A}\Big)\Big| \sqrt{4LA^2 + (1+4L)o(A^2) \sum_{i=1}^{2\lfloor\beta\rfloor+1} c'_{(i;\beta)} \Big|\frac{\lambda\pi}{A}\Big|^i + c'_{(2\lceil\beta\rceil;\beta)} \Big|\frac{\lambda\pi}{A}\Big|^{2\lceil\beta\rceil} o(A^2)} \right\}^2 \\
\leq &\left\{ \frac{1}{2A} \sum_{\lambda\in\mathbb{Z}} \Big| \widehat{f_0 * g_A}\Big(\frac{\lambda\pi}{A}\Big)\Big| \left[\sqrt{4LA^2} + \sqrt{(1+4L)o(A^2)} \sum_{i=1}^{2\lfloor\beta\rfloor+1} \sqrt{c'_{(i;\beta)}} \Big|\frac{\lambda\pi}{A}\Big|^{i/2} \right.\right. \\
&\left.\left. + \sqrt{c'_{(2\lceil\beta\rceil;\beta)}} \Big|\frac{\lambda\pi}{A}\Big|^{\lceil\beta\rceil} o(A) \right] \right\}^2 \\
= &\left\{ \sqrt{4LA^2} \frac{1}{2A} \sum_{\lambda\in\mathbb{Z}} \Big| \widehat{f_0 * g_A}\Big(\frac{\lambda\pi}{A}\Big)\Big| + \sqrt{(1+4L)o(A^2)} \sum_{i=1}^{2\lfloor\beta\rfloor+1} \sqrt{c'_{(i;\beta)}} \right. \\
&\left. \times \frac{1}{2A} \sum_{\lambda\in\mathbb{Z}} \Big| \Big(\frac{\lambda\pi}{A}\Big)^{i/2} \widehat{f_0 * g_A}\Big(\frac{\lambda\pi}{A}\Big)\Big| + \sqrt{c'_{(2\lceil\beta\rceil;\beta)}} \frac{1}{2A} \sum_{\lambda\in\mathbb{Z}} \Big| \Big(\frac{\lambda\pi}{A}\Big)^{\lceil\beta\rceil} \widehat{f_0 * g_A}\Big(\frac{\lambda\pi}{A}\Big)\Big| o(A) \right\}^2 \qquad (23)
\end{aligned}
$$

By virtue of

$$
\begin{aligned}
\frac{1}{2A} \sum_{\lambda\in\mathbb{Z}} \Big| \widehat{f_0 * g_A}\Big(\frac{\lambda\pi}{A}\Big)\Big| d\sigma &\leq \frac{1+o(1)}{2A-1} \qquad \text{and} \\
\frac{1}{2A} \sum_{\lambda\in\mathbb{Z}} \Big| \Big(\frac{\lambda\pi}{A}\Big)^{i/2} \widehat{f_0 * g_A}\Big(\frac{\lambda\pi}{A}\Big)\Big| &\leq \frac{1}{2A-1} \Big\| \widehat{f_0^{(i/2)}} \Big\|_{\ell_1} \qquad \text{for all } i = 1, \ldots, 2\lceil\beta\rceil \qquad (24)
\end{aligned}
$$

which we prove in the sequel, we can proceed at inequality number (23) by:

$$
\begin{aligned}
&\Big\| \Big(f_0 * g_A \cdot \sum_{k\neq 0} \theta_k \varphi_k \Big)^{(\beta)} \Big\|_2^2 \\
\leq &\left\{ \sqrt{4LA^2} \frac{1+o(1)}{2A-1} + \sqrt{(1+4L)o(A^2)} \sum_{i=1}^{2\lfloor\beta\rfloor+1} \sqrt{c'_{(i;\beta)}} \frac{1}{2A-1} \Big\| \widehat{f_0^{(i/2)}} \Big\|_{\ell_1} \right. \\
&\left. + \sqrt{c'_{(2\lceil\beta\rceil;\beta)}} \frac{1}{2A-1} \Big\| \widehat{f_0^{(\lceil\beta\rceil)}} \Big\|_{\ell_1} o(A) \right\}^2 \\
= &\Big\{ \sqrt{L} + o(1) \Big\}^2 \Big(1 + o(1) \Big) \\
= &L + o(1) \qquad (25)
\end{aligned}
$$

which is the desired result. For (24) we proceed in a way similar to the one used to derive (21). For $\gamma = 0, \frac{1}{2}, \ldots, \lfloor\beta\rfloor + \frac{1}{2}, \lceil\beta\rceil$:

$$
\begin{aligned}
\frac{1}{2A} \sum_{\lambda\in\mathbb{Z}} \Big| \Big(\frac{\lambda\pi}{A}\Big)^{\gamma} \widehat{f_0 * g_A}\Big(\frac{\lambda\pi}{A}\Big)\Big| &= \frac{1}{2A} \sum_{\lambda\in\mathbb{Z}} \Big| \Big(\frac{\lambda\pi}{A}\Big)^{\gamma} \widehat{f}_0\Big(\frac{\lambda\pi}{A}\Big) \frac{2\sin\frac{(A-1/2)\lambda\pi}{A}}{(2A-1)\frac{\lambda\pi}{A}} \Big| \\
&= \frac{1}{2A-1} \frac{1}{2A} \sum_{\lambda\in\mathbb{Z}} \Big| \Big(\frac{\lambda\pi}{A}\Big)^{\gamma} \widehat{f}_0\Big(\frac{\lambda\pi}{A}\Big) \frac{\sin\frac{\lambda\pi}{2A}}{\frac{\lambda\pi}{2A}} \Big|
\end{aligned}
$$

As mentioned at the beginning of the section, f_0 was selected so as to satisfy $\int |\widehat{f}_0(\omega)\frac{\sin\omega/2}{\omega/2}|d\omega = 2\pi$, such that

$$\frac{1}{2A}\sum_{\lambda\in\mathbb{Z}}\Big|\widehat{f}_0\Big(\frac{\lambda\pi}{A}\Big)\ \frac{\sin\frac{\lambda\pi}{2A}}{\frac{\lambda\pi}{2A}}\Big| \ = \ \frac{1}{2\pi}\int\Big|\widehat{f}_0(\omega)\ \frac{\sin\,\omega/2}{\omega/2}\Big|d\omega\ \Big(1+o(1)\Big) \ = \ 1+o(1),$$

Moreover it holds that

$$\frac{1}{2A}\sum_{\lambda\in\mathbb{Z}}\Big|\Big(\frac{\lambda\pi}{A}\Big)^{\gamma}\widehat{f}_0\Big(\frac{\lambda\pi}{A}\Big)\ \frac{\sin\frac{\lambda\pi}{2A}}{\frac{\lambda\pi}{2A}}\Big| \ \leq\ \frac{1}{2A}\sum_{\lambda\in\mathbb{Z}}\Big|\Big(\frac{\lambda\pi}{A}\Big)^{\gamma}\widehat{f}_0\Big(\frac{\lambda\pi}{A}\Big)\Big| \ = \ \Big\|\widehat{f_0^{(\gamma)}}\Big\|_{\ell_1} \ \leq\ \text{const.}\ <\ \infty$$

because $\|\widehat{f_0^{(\gamma)}}\|_{\ell_1} = \frac{1}{2\pi}\int|\omega^{\gamma}\widehat{f}_0(\omega)|d\omega\,(1+o(1))$ exists or all γ. Hence we have obtained (24). In connection with (19), (21), and (25) the proof of Theorem III.3.1 is completed:

$$\begin{aligned}
\|f_\theta^{(\beta)}\|_2 \ &\leq\ \frac{1}{b(\theta)}\Big\|(f_0 * g_A)^{(\beta)}\Big\|_2 + \frac{1}{b(\theta)}\Big\|\Big(f_0 * g_A\sum_{k\neq 0}\theta_k\varphi_k\Big)^{(\beta)}\Big\|_2 \\
&=\ O\left(A^{-1}\right)\|f_0^{(\beta)}\|_2 + \sqrt{L}\Big(1+o(1)\Big) \\
&=\ \sqrt{L}+o(1)
\end{aligned}$$

□

Lemma III.3.5 There exist appropriate constants $c'_{(i;\beta)} < \infty$, such that for $\theta \in \Theta_A(L)$

$$\begin{aligned}
\frac{1}{2A}\sum_{\kappa\in\mathbb{Z}}\Big|\Big(\frac{\kappa\pi}{A}\Big)^{\beta}\sum_{k\neq 0}\theta_k\widehat{\varphi}_k\Big(\frac{(\kappa-\lambda)\pi}{A}\Big)\Big|^2 \ &= \ 4LA^2 + (1+4L)o(A^2)\sum_{i=1}^{2\lfloor\beta\rfloor+1} c'_{(i;\beta)}\Big|\frac{\lambda\pi}{A}\Big|^i \\
&\quad + c'_{(2\lceil\beta\rceil;\beta)}\Big|\frac{\lambda\pi}{A}\Big|^{2\lceil\beta\rceil}o(A^2)
\end{aligned}$$

Proof The power function is an analytical function. We may thus expand $(\frac{\kappa\pi}{A})^{\beta}$ into an infinite Taylor series at point $\frac{(\kappa-\lambda)\pi}{A}$.

$$\Big(\frac{\kappa\pi}{A}\Big)^{\beta} \ = \ \sum_{i=0}^{\infty}\frac{\beta\cdots(\beta-i+1)}{i!}\ \Big(\frac{\lambda\pi}{A}\Big)^{i}\ \Big(\frac{(\kappa-\lambda)\pi}{A}\Big)^{\beta-i}$$

Recall the definition: $\lfloor x\rfloor$ is the integer part of a real number x, and for x positive (as in our case) $\lceil x\rceil := \lfloor x\rfloor + 1$. We cut the Tailor expansion of $(\frac{\kappa\pi}{A})^{\beta}$ after $\lfloor\beta\rfloor$ and bound the residual.

$$\begin{aligned}
&\Big|\sum_{i=\lceil\beta\rceil}^{\infty}\frac{\beta\cdots(\beta-i+1)}{i!}\ \Big(\frac{\lambda\pi}{A}\Big)^{i}\Big(\frac{(\kappa-\lambda)\pi}{A}\Big)^{\beta-i}\Big| \\
=\ &\Big|\sum_{i=0}^{\infty}\frac{\beta\cdots(\beta-(\lceil\beta\rceil+i)+1)}{(\lceil\beta\rceil+i)!}\ \Big(\frac{\lambda\pi}{A}\Big)^{\lceil\beta\rceil+i}\Big(\frac{(\kappa-\lambda)\pi}{A}\Big)^{\beta-(\lceil\beta\rceil+i)}\Big| \\
\leq\ &\frac{\beta\cdots(\beta-\lfloor\beta\rfloor)}{\lceil\beta\rceil!}\ \Big|\frac{\lambda\pi}{A}\Big|^{\lceil\beta\rceil}\sum_{i=0}^{\infty}\Big|\frac{(\beta-\lceil\beta\rceil)\cdots(\beta-\lfloor\beta\rfloor-i)}{i!}\Big|\ \Big|\frac{\lambda\pi}{A}\Big|^{i}\Big|\frac{(\kappa-\lambda)\pi}{A}\Big|^{\beta-\lceil\beta\rceil-i} \\
=\ &\frac{\beta\cdots(\beta-\lfloor\beta\rfloor)}{\lceil\beta\rceil!}\ \Big|\frac{\lambda\pi}{A}\Big|^{\lceil\beta\rceil}\sum_{i=0}^{\infty}\frac{(-1)^i(\beta-\lceil\beta\rceil)\cdots(\beta-\lfloor\beta\rfloor-i)}{i!}\ \Big|\frac{\lambda\pi}{A}\Big|^{i}\Big|\frac{(\kappa-\lambda)\pi}{A}\Big|^{\beta-\lceil\beta\rceil-i}
\end{aligned}$$

$$\begin{aligned}
&= \frac{\beta\cdots(\beta-\lfloor\beta\rfloor)}{\lceil\beta\rceil!}\ \Big|\frac{\lambda\pi}{A}\Big|^{\lceil\beta\rceil}\sum_{i=0}^{\infty}\frac{(\beta-\lceil\beta\rceil)\cdots(\beta-\lfloor\beta\rfloor-i)}{i!}\ \Big(\frac{-|\lambda|\pi}{A}\Big)^{i}\Big(\frac{|\kappa-\lambda|\pi}{A}\Big)^{\beta-\lceil\beta\rceil-i}\\
&= \frac{\beta\cdots(\beta-\lfloor\beta\rfloor)}{\lceil\beta\rceil!}\ \Big|\frac{\lambda\pi}{A}\Big|^{\lceil\beta\rceil}\Big(\frac{(|\kappa-\lambda|-|\lambda|)\pi}{A}\Big)^{\beta-\lceil\beta\rceil}\\
&=: c_{(\lceil\beta\rceil;\beta)}\ \Big|\frac{\lambda\pi}{A}\Big|^{\lceil\beta\rceil}\Big(\frac{(|\kappa-\lambda|-|\lambda|)\pi}{A}\Big)^{\beta-\lceil\beta\rceil}
\end{aligned}$$

such that we have

$$\begin{aligned}
\Big|\frac{\kappa\pi}{A}\Big|^{\beta} &\leq \Big|\sum_{i=0}^{\lfloor\beta\rfloor}\frac{\beta\cdots(\beta-i+1)}{i!}\ \Big(\frac{\lambda\pi}{A}\Big)^{i}\Big(\frac{(\kappa-\lambda)\pi}{A}\Big)^{\beta-i}\Big| + c_{(\lceil\beta\rceil;\beta)}\ \Big|\frac{\lambda\pi}{A}\Big|^{\lceil\beta\rceil}\Big(\frac{(|\kappa-\lambda|-|\lambda|)\pi}{A}\Big)^{\beta-\lceil\beta\rceil}\\
&=: \Big|\sum_{i=0}^{\lfloor\beta\rfloor}c_{(i;\beta)}\Big(\frac{\lambda\pi}{A}\Big)^{i}\Big(\frac{(\kappa-\lambda)\pi}{A}\Big)^{\beta-i}\Big| + c_{(\lceil\beta\rceil;\beta)}\ \Big|\frac{\lambda\pi}{A}\Big|^{\lceil\beta\rceil}\Big(\frac{(|\kappa-\lambda|-|\lambda|)\pi}{A}\Big)^{\beta-\lceil\beta\rceil} \qquad (26)
\end{aligned}$$

When $\beta \in \mathbb{N}$, then $c_{(\lceil\beta\rceil;\beta)}$ is equal to 0. Because κ and $\lambda \in \mathbb{N}$, it holds for $\beta \in \mathbb{R}^+\backslash\mathbb{N}$, that $(\frac{(|\kappa-\lambda|-|\lambda|)\pi}{A})^{\beta-\lceil\beta\rceil} = O(A^{\lceil\beta\rceil-\beta}) = o(A)$. By squaring the right-hand side of inequality (26), we obtain

$$\begin{aligned}
&\left(\Big|\sum_{i=0}^{\lfloor\beta\rfloor}c_{(i;\beta)}\Big(\frac{\lambda\pi}{A}\Big)\Big(\frac{(\kappa-\lambda)\pi}{A}\Big)^{\beta-i}\Big| + c_{(\lceil\beta\rceil;\beta)}\ \Big|\frac{\lambda\pi}{A}\Big|^{\lceil\beta\rceil}\Big(\frac{(|\kappa-\lambda|-|\lambda|)\pi}{A}\Big)^{\beta-\lceil\beta\rceil}\right)^2\\
&= \sum_{i,j=0}^{\lfloor\beta\rfloor}c_{(i;\beta)}c_{(i;\beta)}\Big|\frac{\lambda\pi}{A}\Big|^{i+j}\Big|\frac{(\kappa-\lambda)\pi}{A}\Big|^{2\beta-i-j} + 2\,c_{(\lceil\beta\rceil;\beta)}o(A)\sum_{i=0}^{\lfloor\beta\rfloor}c_{(i;\beta)}\Big|\frac{\lambda\pi}{A}\Big|^{\lceil\beta\rceil+i}\Big|\frac{(\kappa-\lambda)\pi}{A}\Big|^{\beta-i}\\
&\quad + c^2_{(\lceil\beta\rceil;\beta)}\ \Big|\frac{\lambda\pi}{A}\Big|^{2\lceil\beta\rceil}o(A^2)\\
&=: \Big|\frac{(\kappa-\lambda)\pi}{A}\Big|^{2\beta} + o(A)\sum_{i=1}^{2\lfloor\beta\rfloor+1}c'_{(i;\beta)}\Big|\frac{\lambda\pi}{A}\Big|^{i}\Big|\frac{(\kappa-\lambda)\pi}{A}\Big|^{2\beta-i} + c'_{(\lceil\beta\rceil;\beta)}\ \Big|\frac{\lambda\pi}{A}\Big|^{2\lceil\beta\rceil}o(A^2)
\end{aligned}$$

with appropriate constants $c'_{(i;\beta)}$. Plugging this Taylor series expansion into the sum in question, we are led to:

$$\begin{aligned}
&\frac{1}{2A}\sum_{\kappa\in\mathbb{Z}}\Big|\Big(\frac{\kappa\pi}{A}\Big)^{\beta}\sum_{k\neq 0}\theta_k\widehat{\varphi}_k\Big(\frac{(\kappa-\lambda)\pi}{A}\Big)\Big|^2\\
&= \frac{1}{2A}\sum_{\kappa\in\mathbb{Z}}\left(\Big|\frac{(\kappa-\lambda)\pi}{A}\Big|^{2\beta} + o(A)\sum_{i=1}^{2\lfloor\beta\rfloor+1}c'_{(i;\beta)}\Big|\frac{\lambda\pi}{A}\Big|^{i}\Big|\frac{(\kappa-\lambda)\pi}{A}\Big|^{2\beta-i} + c'_{(\lceil\beta\rceil;\beta)}\ \Big|\frac{\lambda\pi}{A}\Big|^{2\lceil\beta\rceil}o(A^2)\right)\\
&\qquad \times\Big|\sum_{k\neq 0}\theta_k\widehat{\varphi}_k\Big(\frac{(\kappa-\lambda)\pi}{A}\Big)\Big|^2\\
&= \frac{1}{2A}\sum_{\kappa\in\mathbb{Z}}\left(\Big|\Big(\frac{(\kappa-\lambda)\pi}{A}\Big)^{\beta}\sum_{k\neq 0}\theta_k\widehat{\varphi}_k\Big(\frac{(\kappa-\lambda)\pi}{A}\Big)\Big|^2 + o(A)\sum_{i=1}^{2\lfloor\beta\rfloor+1}c'_{(i;\beta)}\Big|\frac{\lambda\pi}{A}\Big|^{i}\Big|\Big(\frac{(\kappa-\lambda)\pi}{A}\Big)^{\beta-i/2}\right.\\
&\qquad \left.\times\sum_{k\neq 0}\theta_k\widehat{\varphi}_k\Big(\frac{(\kappa-\lambda)\pi}{A}\Big)\Big|^2 + c'_{(\lceil\beta\rceil;\beta)}\ \Big|\frac{\lambda\pi}{A}\Big|^{2\lceil\beta\rceil}\Big|\sum_{k\neq 0}\theta_k\widehat{\varphi}_k\Big(\frac{(\kappa-\lambda)\pi}{A}\Big)\Big|^2 o(A^2)\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2A}\sum_{\kappa\in\mathbb{Z}}\Big|\sum_{k\neq 0}\theta_k\widehat{\varphi_k^{(\beta)}}\Big(\frac{(\kappa-\lambda)\pi}{A}\Big)\Big|^2 + o(A)\sum_{i=1}^{2\lfloor\beta\rfloor+1} c'_{(i;\beta)}\Big|\frac{\lambda\pi}{A}\Big|^i \frac{1}{2A}\sum_{\kappa\in\mathbb{Z}}\Big|\sum_{k\neq 0}\theta_k\widehat{\varphi_k^{(\beta-i/2)}}\Big(\frac{(\kappa-\lambda)\pi}{A}\Big)\Big|^2 \\
&\quad + c'_{(\lceil\beta\rceil;\beta)}\Big|\frac{\lambda\pi}{A}\Big|^{2\lceil\beta\rceil}\frac{1}{2A}\sum_{\kappa\in\mathbb{Z}}\Big|\sum_{k\neq 0}\theta_k\widehat{\varphi}_k\Big(\frac{(\kappa-\lambda)\pi}{A}\Big)\Big|^2 o(A^2) \\
&= \Big\|\sum_{k\neq 0}\theta_k\widehat{\varphi_k^{(\beta)}}\Big\|_{\ell_2}^2 + o(A)\sum_{i=1}^{2\lfloor\beta\rfloor+1} c'_{(i;\beta)}\Big|\frac{\lambda\pi}{A}\Big|^i\Big\|\sum_{k\neq 0}\theta_k\widehat{\varphi_k^{(\beta-i/2)}}\Big\|_{\ell_2}^2 + c'_{(\lceil\beta\rceil;\beta)}\Big|\frac{\lambda\pi}{A}\Big|^{2\lceil\beta\rceil}\Big\|\sum_{k\neq 0}\theta_k\widehat{\varphi}_k\Big\|_{\ell_2}^2 o(A^2) \\
&= \Big\|\sum_{k\neq 0}\theta_k\varphi_k^{(\beta)}\Big\|_2^2 + o(A)\sum_{i=1}^{2\lfloor\beta\rfloor+1} c'_{(i;\beta)}\Big|\frac{\lambda\pi}{A}\Big|^i\Big\|\sum_{k\neq 0}\theta_k\varphi_k^{(\beta-i/2)}\Big\|_2^2 \\
&\quad + c'_{(\lceil\beta\rceil;\beta)}\Big|\frac{\lambda\pi}{A}\Big|^{2\lceil\beta\rceil}\Big\|\sum_{k\neq 0}\theta_k\varphi_k\Big\|_2^2 o(A^2) \qquad (27)
\end{aligned}
$$

In the next step we employ:

$$
\Big\|\sum_{k\neq 0}\theta_k\varphi_k^{(\gamma)}\Big\|_2^2 = \sum_{k\neq 0}\theta_k^2\left(\frac{k\pi}{A}\right)^{2\gamma} \qquad (28)
$$

proven in (30). Furthermore for $\Theta_A(L)$, $\sum|\theta_k| \leq A^{-2\beta+1}$ and $\sum\theta_k^2\left(\frac{k\pi}{A}\right)^{2\beta} \leq 4A^2L$ had been determined in (7). Therefrom we can in (31) show that

$$
\sum_{k\neq 0}\theta_k^2\left(\frac{k\pi}{A}\right)^{2\beta-l} \leq (1+4L)A^{2-l} \qquad \text{for } \; 0<l<2\beta \qquad (29)
$$

And thus

$$
\begin{aligned}
\frac{1}{2A}\sum_{\kappa\in\mathbb{Z}}\Big|\Big(\frac{\kappa\pi}{A}\Big)^\beta\sum_{k\neq 0}\theta_k\widehat{\varphi}_k\Big(\frac{(\kappa-\lambda)\pi}{A}\Big)\Big|^2 &= 4LA^2 + o(A)\sum_{i=1}^{2\lfloor\beta\rfloor+1} c'_{(i;\beta)}\Big|\frac{\lambda\pi}{A}\Big|^i(1+4L)A^{2-i} \\
&\quad + c'_{(\lceil\beta\rceil;\beta)}\Big|\frac{\lambda\pi}{A}\Big|^{2\lceil\beta\rceil}A^{-2\beta+1}o(A^2) \\
&= 4LA^2 + (1+4L)o(A^2)\sum_{i=1}^{2\lfloor\beta\rfloor+1} c'_{(i;\beta)}\Big|\frac{\lambda\pi}{A}\Big|^i \\
&\quad + c'_{(\lceil\beta\rceil;\beta)}\Big|\frac{\lambda\pi}{A}\Big|^{2\lceil\beta\rceil}o(A^2)
\end{aligned}
$$

which is the statement of Lemma III.3.5. We are now left to verify the intermediate steps (28) and (29).
As an exception to the ordinary case, sine and cosine enjoy an easy to calculate fractional derivative: $\sin^{(\gamma)}(ax) = a^\gamma\sin\,(ax+\gamma\pi/2)$ and the like for cosine (Samko (1987), p. 174). Obviously, the orthogonality between our functions φ_k is preserved through derivation.

$$
\begin{aligned}
&\int\Big(\sum_{k\neq 0}\theta_k\varphi_k^{(\gamma)}(x)\Big)^2 dx \\
&= \int\sum_{k\neq 0}\theta_k^2\;\varphi_k^{(\gamma)}(x)^2 dx
\end{aligned}
$$

$$
\begin{aligned}
&= \int_{-A}^{A} \sum_{k>0} \theta_k^2 \, \frac{1}{A} \left(\frac{k\pi}{A}\right)^{2\gamma} \cos^2\left(\frac{k\pi x}{A} + \frac{\gamma\pi}{2}\right) + \sum_{k<0} \theta_k^2 \, \frac{1}{A} \left(\frac{k\pi}{A}\right)^{2\gamma} \sin^2\left(\frac{k\pi x}{A} + \frac{\gamma\pi}{2}\right) \, dx \\
&= \sum_{k>0} \theta_k^2 \, \left(\frac{k\pi}{A}\right)^{2\gamma} \frac{1}{A} \int_{-A}^{A} \cos^2\left(\frac{k\pi x}{A} + \frac{\gamma\pi}{2}\right) + \sum_{k<0} \theta_k^2 \left(\frac{k\pi}{A}\right)^{2\gamma} \frac{1}{A} \int_{-A}^{A} \sin^2\left(\frac{k\pi x}{A} + \frac{\gamma\pi}{2}\right) \, dx \\
&= \sum_{k>0} \theta_k^2 \, \left(\frac{k\pi}{A}\right)^{2\gamma} \frac{1}{A} \int_{-A}^{A} \cos^2 \frac{k\pi x}{A} \, dx + \sum_{k<0} \theta_k^2 \left(\frac{k\pi}{A}\right)^{2\gamma} \frac{1}{A} \int_{-A}^{A} \sin^2 \frac{k\pi x}{A} \, dx \\
&= \sum_{k>0} \theta_k^2 \, \left(\frac{k\pi}{A}\right)^{2\gamma} \frac{1}{A} \, A + \sum_{k<0} \theta_k^2 \left(\frac{k\pi}{A}\right)^{2\gamma} \frac{1}{A} \, A \\
&= \sum_{k\neq 0} \theta_k^2 \, \left(\frac{k\pi}{A}\right)^{2\gamma} \qquad (30)
\end{aligned}
$$

Referring to step (29):

$$
\begin{aligned}
\sum_{k\neq 0} \theta_k^2 \left(\frac{k\pi}{A}\right)^{2\beta-l} &= \sum_{0\neq|k|\leq A^2/\pi} \theta_k^2 \left(\frac{k\pi}{A}\right)^{2\beta-l} + \sum_{|k|>A^2/\pi} \theta_k^2 \left(\frac{k\pi}{A}\right)^{2\beta} \left(\frac{k\pi}{A}\right)^{-l} \\
&\leq A^{2\beta-l} \sum_{0\neq|k|\leq A^2/\pi} \theta_k^2 + A^{-l} \sum_{|k|>A^2/\pi} \theta_k^2 \left(\frac{k\pi}{A}\right)^{2\beta} \\
&\leq A^{2\beta-l} \sum_{k\neq 0} |\theta_k| + A^{-l} \sum_{k\neq 0} \theta_k^2 \left(\frac{k\pi}{A}\right)^{2\beta} \\
&\leq A^{2\beta-l} \cdot A^{-2\beta+1} + A^{-l} \cdot 4LA^2 \\
&= A^{1-l} + 4LA^{2-l} \\
&\leq (1+4L)A^{2-l} \qquad (31)
\end{aligned}
$$

This concludes the proof of Lemma III.3.5. □

Proof of **Lemma III.3.4** We start with

$$
\begin{aligned}
I_{f_\theta}(\theta_\kappa) &= \int_{-A}^{A} \frac{(\partial f_\theta(x)/\partial\theta_\kappa)^2}{f_\theta(x)} \, dx \\
&= \frac{1}{b^2(\theta)} \int_{-A}^{A} \frac{1}{f_\theta(x)} \left[-f_\theta(x) \int_{-A}^{A} f_0 * g_A(y)\varphi_\kappa(y) \, dy + f_0 * g_A(x)\varphi_\kappa(x) \right]^2 dx \\
&= \frac{1}{b^2(\theta)} \int_{-A}^{A} \Bigg[f_\theta(x) \left(\int_{-A}^{A} f_0 * g_A(y)\varphi_\kappa(y) \, dy \right)^2 - 2 \int_{-A}^{A} f_0 * g_A(y)\varphi_\kappa(y) \, dy \; f_0 * g_A(x)\varphi_\kappa(x) \\
&\qquad + \frac{1}{f_\theta(x)} \Big(f_0 * g_A(x)\varphi_\kappa(x) \Big)^2 \Bigg] \, dx \\
&= -\frac{1}{b^2(\theta)} \left[\int_{-A}^{A} f_0 * g_A(y)\varphi_\kappa(y) \, dy \right]^2 + \frac{1}{b^2(\theta)} \int_{-A}^{A} \frac{\Big(f_0 * g_A(x) \Big)^2 \varphi_\kappa^2(x)}{\frac{1}{b(\theta)} f_0 * g_A(x) \Big(1 + \sum_{\lambda\neq 0} \theta_\lambda \varphi_\lambda(x) \Big)} dx \\
&= -\frac{1}{b^2(\theta)} \left[\int_{-A}^{A} f_0 * g_A(y)\varphi_\kappa(y) \, dy \right]^2 + \frac{1}{b(\theta)} \int_{-A}^{A} \frac{f_0 * g_A(x)\varphi_\kappa^2(x)}{1 + \sum_{\lambda\neq 0} \theta_\lambda \varphi_\lambda(x)} \, dx
\end{aligned}
$$

$$
\begin{aligned}
= & \; -\frac{1}{b^2(\theta)}\left[\int_{-A}^{A}\frac{\varphi_\kappa(x)}{2A-1}\,dx - 2\int_{A-1}^{A}\left(\frac{1}{2A-1}-f_0*g_A(x)\right)\varphi_\kappa(x)\,dx\right]^2 \\
& + \frac{1}{b(\theta)}\left[\int_{-A}^{A}\frac{\frac{1}{2A-1}\,\varphi_\kappa^2(x)}{1+\sum\limits_{\lambda\neq 0}\theta_\lambda\varphi_\lambda(x)}\,dx - 2\int_{A-1}^{A}\frac{\left(\frac{1}{2A-1}-f_0*g_A(x)\right)\varphi_\kappa^2(x)}{1+\sum\limits_{\lambda\neq 0}\theta_\lambda\varphi_\lambda(x)}\,dx\right]
\end{aligned}
$$

The leading term is $\int_{-A}^{A}\frac{\frac{1}{2A-1}\,\varphi_\kappa^2(x)}{1+\sum\theta_\lambda\varphi_\lambda(x)}\,dx$. Due to $|\sum\theta_\lambda\varphi_\lambda(x)|\le A^{-1/2}\sum|\theta_\lambda|\le A^{-1/2}\cdot A^{-2\beta+1} < A^{-1/2}$, we know it lies in the interval $((2A-1)^{-1}(1+A^{-1/2})^{-1},(2A-1)^{-1}(1-A^{-1/2})^{-1})$. Moreover from (19) we have $1-A^{-3/2}\le b(\theta)\le 1+A^{-3/2}$. For $A\longrightarrow\infty$ we obtain:

$$
\begin{aligned}
I_{f_\theta}(\theta_\kappa) &= \frac{1}{b^2(\theta)}\left[0+O\left(\frac{1}{(A-1/2)\sqrt{A}}\right)\right]^2+\frac{1}{b(\theta)}\left[\frac{1+o(1)}{2A-1}+O\left(\frac{1+o(1)}{(A-1/2)A}\right)\right] \\
&= \Big(1+o(1)\Big)O\left(A^{-3}\right)+\Big(1+o(1)\Big)\left[\frac{1+o(1)}{2A-1}+O\left(A^{-2}\right)\right] \\
&= \frac{1+o(1)}{2A}
\end{aligned}
$$

$\mathrm{Re}\widehat{f}_\theta\left(\frac{\kappa\pi}{A}\right)$ can be expressed as $A^{1/2}\int f_\theta(x)\varphi_{|\kappa|}(x)dx$, yielding

$$
\begin{aligned}
\frac{\partial\,\mathrm{Re}\widehat{f}_\theta\left(\frac{\kappa\pi}{A}\right)}{\partial\theta_{|\kappa|}} &= \frac{\partial}{\partial\theta_{|\kappa|}}\sqrt{A}\int_{-A}^{A}f_\theta(x)\varphi_{|\kappa|}(x)dx \\
&= \frac{\sqrt{A}}{b(\theta)}\left[-\int_{-A}^{A}f_\theta(x)\int_{-A}^{A}f_0*g_A(y)\varphi_{|\kappa|}(y)dy\;\varphi_{|\kappa|}(x)dx+\int_{-A}^{A}f_0*g_A(x)\varphi_{|\kappa|}^2(x)dx\right] \\
&= \frac{\sqrt{A}}{b(\theta)}\left[-\frac{\mathrm{Re}\widehat{f}_\theta\left(\frac{\kappa\pi}{A}\right)}{\sqrt{A}}\int_{-A}^{A}f_0*g_A(y)\varphi_{|\kappa|}(y)dy+\int_{-A}^{A}f_0*g_A(x)\varphi_{|\kappa|}^2(x)dx\right] \\
&= \frac{\mathrm{Re}\widehat{f}_\theta\left(\frac{\kappa\pi}{A}\right)}{b(\theta)}\left[-\int_{-A}^{A}\frac{1}{2A-1}\,\varphi_{|\kappa|}(y)dy+2\int_{A-1}^{A}\left(\frac{1}{2A-1}-f_0*g_A(y)\right)\varphi_{|\kappa|}(y)dy\right] \\
&\quad+\frac{\sqrt{A}}{b(\theta)}\left[\int_{-A}^{A}\frac{1}{2A-1}\,\varphi_{|\kappa|}^2(x)dx-2\int_{A-1}^{A}\left(\frac{1}{2A-1}-f_0*g_A(x)\right)\varphi_{|\kappa|}^2(x)dx\right] \\
&= \frac{\mathrm{Re}\widehat{f}_\theta\left(\frac{\kappa\pi}{A}\right)}{1+o(1)}\left[0+O\left(\frac{1}{(A-1/2)\sqrt{A}}\right)\right]+\frac{\sqrt{A}}{1+o(1)}\left[\frac{1}{2A-1}+O\left(\frac{1}{(A-1/2)A}\right)\right] \\
&= \frac{1+o(1)}{2\sqrt{A}}
\end{aligned}
$$

A similar result is obtained for $\mathrm{Im}\widehat{f}_\theta\left(\frac{\kappa\pi}{A}\right)$, whereby $\int f_0*g_A(y)\varphi_{-|\kappa|}(y)dy=0$, because the sine function is anti-symmetric.

A APPENDIX

Appendix A's destination is but to simplify the reader's reference to results which are essential to definitions and proofs in the present work, Part I through III. There is no claim on authorship neither to propositions nor to demonstrations.

A.1 Fourier transform and expansion on L_2

The following discussion is extracted from Chandrasekharan: Classical Fourier transforms (1989).

A complex-valued function $f(x)$ of the real variable x is said to belong to *Schwartz's space* $\mathcal{S}$, if f is differentiable infinitely often, and for any integers p and q

$$x^p f^{(q)} \longrightarrow 0, \text{ as } |x| \longrightarrow \infty.$$

$\mathcal{S}$ is a dense subset of $L_2(-\infty, \infty)$. For $f \in \mathcal{S}$ we define the Fourier transform $\widehat{f}$ of f by

$$\widehat{f}(\omega) := \int f(x) e^{i\omega x} dx, \; -\infty < \omega < \infty.$$

Theorem: If $f \in \mathcal{S}$, then also $\widehat{f} \in \mathcal{S}$.

Proof: If $f \in \mathcal{S}$, then $\|f^{(q)}\|_1 < \infty$, for all $q \in \mathbb{N}$. Let ω be real, $\omega \neq 0$. For $R > 0$ we have

$$\int_{-R}^{R} f(x) e^{i\omega x} dx = f(x) \frac{e^{i\omega x}}{i\omega} \bigg|_{-R}^{R} - \int_{-R}^{R} f'(x) \frac{e^{i\omega x}}{i\omega} dx.$$

For

$$f(x) - f(0) = \int_0^x f'(t) dt, \quad \text{and} \quad f' \in L_1(-\infty, \infty),$$

so that

$$f(0) + \int_0^\infty f'(t) dt = f(\infty) = 0 \quad \text{and the like for } -\infty.$$

Hence, on letting $R \longrightarrow \infty$, we have

$$\int e^{i\omega x} f(x) dx = -\int f'(x) \frac{e^{i\omega x}}{i\omega} \, dx, \; \omega \neq 0,$$

which implies that

$$-i\omega \; \widehat{f}(\omega) = \widehat{f'}(\omega), \text{ for } \omega \neq 0.$$

This holds also for $\omega = 0$ by continuity (since the left-hand side is zero, while the right-hand side is also $\int f'(x) dx = f(\infty) - f(-\infty) = 0$). Since

$$|\omega \widehat{f}(\omega)| \leq \int |f'(x)| dx = \|f'\|_1 < \infty,$$

we conclude that $\widehat{f}(\omega) = O(1/|\omega|)$.

Now, on the other hand if $f \in \mathcal{S}$, then $\int |x^p f(x)| dx < \infty$ for all $p \in \mathbb{N}$. Let

$$f_h(x) := f(x) \left(\frac{e^{ihx} - 1}{h} \right)$$

for real h, and $h \neq 0$. Then

$$\widehat{f}_h(\omega) = \frac{\widehat{f}(\omega + h) - \widehat{f}(\omega)}{h}.$$

Now $f_h(x) \longrightarrow ix\; f(x)$ pointwisely, for almost every x, as $h \longrightarrow 0$; and

$$|f_h(x)| \leq |f(x)| \cdot \left| \frac{e^{ihx} - 1}{h} \right| \leq |x| \cdot |f(x)|.$$

Hence

$$|f_h(x) - ix\; f(x)| \leq 2|x| \cdot |f(x)|.$$

By the Theorem on dominated convergence, we conclude that $\int |f_h(x) - ix\, f(x)|dx \longrightarrow 0$, as $h \longrightarrow 0$. By

$$\sup_{-\infty<\omega<\infty} |\widehat{f}_h(\omega) - \int ix\; f(x)e^{i\omega x}dx| \leq \int |f_h(x) - ix\; f(x)|dx,$$

it follows that

$$\widehat{f}_h(\omega) \longrightarrow \int ix\; f(x)e^{i\omega x}dx,$$

uniformly, as $h \longrightarrow 0$. The last integral is a bounded, continuous function of ω for $-\infty < \omega < \infty$. However, from $\widehat{f}_h(\omega) = h^{-1}(\widehat{f}(\omega + h) - \widehat{f}(\omega))$, we see that $\lim_{h\to 0} \widehat{f}_h(\omega) = \widehat{f}'(\omega)$, hence

$$\widehat{f}\,'(\omega) = \int ix\; f(x)e^{i\omega x}dx.$$

Induction on q and $p \in \mathbb{N}$ gives a more general form of the representation of the derivatives of f in the Fourier domain, and a formula for the derivatives of $\widehat{f}$, respectively:

$$(-i\omega)^q \widehat{f}(\omega) = \widehat{f^{(q)}}(\omega), \text{ where } |\widehat{f}(\omega)| = O(|\omega|^{-q}), \quad \text{and} \quad \widehat{f}^{(p)}(\omega) = \int (ix)^p f(x)e^{i\omega x}dx.$$

From the combination of both, it thus follows that

$$(-i\omega)^q \widehat{f}^{(p)}(\omega) = \int (ix)^p f^{(q)}(x)e^{i\omega x}dx, \text{ where } |\widehat{f^{(q)}}(\omega)| = O(|\omega|^{-p})$$

for all q, $p \in \mathbb{N}$. And therefore $|\widehat{f^{(q)}}(\omega)| = o(|\omega|^{-p})$ for all q, $p \in \mathbb{N}$, which is the assertion. □

On the other hand, we see that, if f and therefore $\widehat{f}$ are in $\mathcal{S}$, then

$$\frac{1}{2\pi} \int \widehat{f}(\omega)e^{-i\omega x}d\omega = \lim_{R\to\infty} \int \widehat{f}(\omega)e^{-|\omega|/R}e^{-i\omega x}dx = f(x),$$

for every $x \in (-\infty, \infty)$. Together with the theorem we have proven that the Fourier transform is a map from $\mathcal{S}$ onto itself, more precisely a linear map.
The composition formula

$$\int f(x)\widehat{g}(x)dx = \int f(x) \int g(y)e^{iyx}dy\, dx = \int\int f(x)e^{iyx}dx\; g(y)dy = \int \widehat{f}(y)g(y)dy$$

and the fact, that if $f \in \mathcal{S}$, then $f(x)$ is $\frac{1}{2\pi}$ times the Fourier transform of the complex conjugate of $\widehat{f}$, $\overline{\widehat{f}} \in \mathcal{S}$, provide us with Parseval's equality:

$$\|f\|_2^2 = \frac{1}{2\pi}\|\widehat{f}\|_2^2$$

and more generally with Plancherel's theorem

$$\int f(x)\overline{g}(x)dx = \frac{1}{2\pi}\int \widehat{f}(\omega)\overline{\widehat{g}}(\omega)d\omega.$$

Besides, it is easy to see that $\widehat{f * g} = \widehat{f}\cdot\widehat{g}$ and $\widehat{f\cdot g} = \frac{1}{2\pi}\widehat{f} * \widehat{g}$.

We are now able to define the Fourier transform on the whole space L_2:
Since $\mathcal{S}$ is a dense subset of $L_2(-\infty,\infty)$, there exists a sequence $(f_n)_{n\in\mathbb{N}}$ of functions belonging to $\mathcal{S}$, such that $\|f_n - f\|_2 \longrightarrow 0$, as $n \longrightarrow \infty$. Hence $\|f_n - f_m\|_2 \longrightarrow 0$ as $n, m \longrightarrow \infty$, but by Plancherel's theorem and the linearity of the map $f_m \longmapsto \widehat{f}_m$, $\|\widehat{f}_n - \widehat{f}_m\|_2 \longrightarrow 0$, as $n, m \longrightarrow \infty$. Since the space $L_2(-\infty,\infty)$ is complete, there exists a function $\varphi \in L_2(-\infty,\infty)$, such that $\|\widehat{f}_n - \varphi\|_2 \longrightarrow 0$, as $n \longrightarrow \infty$. The function φ is defined almost everywhere on $(-\infty,\infty)$. We call φ the Fourier transform of $f \in L_2(-\infty,\infty)$ and denote it by $\widehat{f}$.
We note that φ does not depend on the approximating sequence (f_n). Thus the Fourier transform $\widehat{f}$ of a function $f \in L_2(-\infty,\infty)$ is the limit, in the L_2-norm, of the sequence $(\widehat{f}_n)_{n\in\mathbb{N}}$ of Fourier transforms of any sequence $(f_n)_{n\in\mathbb{N}}$ of functions belonging to $\mathcal{S}$, such that f_n converges in the L_2-norm to the given function f, as $n \longrightarrow \infty$. $\widehat{f} \in L_2(-\infty,\infty)$.

Let us note further: The Fourier transform is a (linear) map from $L_2(-\infty,\infty)$ onto $L_2(-\infty,\infty)$. The inversion formula $f(x) = \frac{1}{2\pi}\int \widehat{f}(\omega)e^{-i\omega x}d\omega$ holds for almost every $x \in (-\infty,\infty)$. Plancherel's theorem holds on the whole space $L_2(-\infty,\infty)$. Every function $f \in L_2(-\infty,\infty)$ is the Fourier transform of a unique element of $L_2(-\infty,\infty)$. For $f, g \in L_2(-\infty,\infty)$, $\widehat{f * g}(\omega) = \widehat{f}(\omega)\widehat{g}(\omega)$ for almost every $\omega \in (-\infty,\infty)$.
The next two properties only hold for functions $f \in L_2 \cap L_1$: If $f(x)$ and $ix \cdot f(x) \in L_2 \cap L_1$, then $\widehat{f'}(\omega) = \int ix\ f(x)e^{i\omega x}dx$ and if $f(x)$ and $f'(x) \in L_2 \cap L_1$, then $\int f'(x)e^{i\omega x}dx = -i\omega\,\widehat{f}(\omega)$.

Fourier expansion is the counterpart to Fourier transform which applies to periodical functions, or as we use it in the proof of Theorem III.3.3, to functions on a bounded interval that can be viewed as periodical when continued to the whole axis.
Let f be a function defined on the $[0,1]$, $\|f\|_2^2 := \int_0^1 f^2(x)dx < \infty$. If this is the case, f is said to be $\in L_2([0,1])$. Let furthermore $(\varphi_\kappa)_{\kappa\in\mathbb{Z}}$, be a complete orthonormal system of functions, likewise $\in L_2([0,1])$. The set $(e^{i2\pi\kappa x})_{\kappa\in\mathbb{Z}}$, is for instance such a system. This time define

$$\widehat{f}(2\pi\kappa) := \int_0^1 f(x)e^{i2\pi\kappa x}dx,$$

the Fourier coefficients of f. Then $(\widehat{f}(2\pi\kappa))_{\kappa\in\mathbb{Z}}$, is called the Fourier expansion of f. It holds that

$$f(x) = \sum_{\kappa\in\mathbb{Z}}\widehat{f}(2\pi\kappa)e^{-i2\pi\kappa x} \quad \text{and} \quad \|f\|_2^2 = \sum_{\kappa\in\mathbb{Z}}\widehat{f}^2(2\pi\kappa).$$

On the other hand, to any given series $(\widehat{f}(2\pi\kappa))_{\kappa\in\mathbb{Z}}$, with $\sum_{\kappa\in\mathbb{Z}}\widehat{f}^2(2\pi\kappa) < \infty$, there exists a function $f \in L_2([0,1])$ of which $(\widehat{f}(2\pi\kappa))_{\kappa\in\mathbb{Z}}$ is the Fourier expansion.
Rules for calculation are similar to the Fourier transform: For f and $g \in L_2([0,1])$, $\widehat{f * g}(2\pi\kappa) = \widehat{f}(2\pi\kappa)\widehat{g}(2\pi\kappa)$, $\widehat{f\cdot g}(2\pi\kappa) = \sum_{\lambda\in\mathbb{Z}}\widehat{f}(2\pi(\kappa-\lambda))\widehat{g}(2\pi\lambda)$. If f is in $L_2([0,1]) \cap L_1([0,1])$, continuously differentiable and f' again $\in L_2([0,1]) \cap L_1([0,1])$, then $\widehat{f'}(2\pi\kappa) = -i2\pi\kappa\,\widehat{f}(2\pi\kappa)$. Evidently, the Fourier expansion can also be carried out on intervals of the real line other than $[0,1]$. Attention must be kept on the orthonormality and completeness of the system of functions defining the Fourier expansion.

A.2 Bernstein type inequality for degenerate U-statistics

As Bernstein's inequality in the case of a sum of independent random variables, it is as well possible to find an exponential bound for the tail probability of a U-statistic.
Let $X_1, \ldots, X_n$ be i.i.d. random variables in $\mathbb{R}$ with common distribution function P. Let $\mathcal{F}$ be a class of measurable real functions on $\mathbb{R}^m$. A U-statistic of order m with kernel $f \in \mathcal{F}$ is defined as

$$U_n^m(f, P) \;:=\; U_n^m(f) \;:=\; \frac{(n-m)!}{n!} \sum_{\mathcal{I}_n^m} f(X_{i_1}, \ldots, X_{i_m})$$

where $\mathcal{I}_n^m$ is the set of all m-tuples $(i_1, \ldots, i_m)$, $1 \leq i_j \leq n$, $i_j \neq i_k$ whenever $j \neq k$, and the U-process indexed by $\mathcal{F}$ is $\{U_n^m(f, P) | f \in \mathcal{F}\}$. Nolan and Pollard (1987, 1988) study the law of large numbers and the central limit theorem for U-processes of order $m = 2$ in terms of chaining inequalities, thereby paralleling the analysis of empirical processes. As an example of the application of their theory, they derive a result, that was already mentioned in the introduction of Part I: asymptotic optimality of bandwidth cross-validation for kernel functions with bounded variation.
Recently discovered, "decoupling"- and "randomization"-techniques (de la Peña (1992) and de la Peña, Montgomery-Smith (1995)) have proven to be very powerful tools in the study of limit theory of more general U-statistics and U-processes, law of large numbers and central limit theorem by Arcones, Giné (1993), law of iterated logarithm by Arcones, Giné (1995). The Bernstein type inequality for degenerate U-statistics is but one achievement brought about by decoupling and randomization.
Since it is sufficient for our purpose, we take advantage of the opportunity to avoid a more complicated notation: in the following considerations, we restrict ourselves to U-statistics of second order and leave higher order U-statistics as well as U-processes aside. Nevertheless, it is possible to prove all results in terms of U-processes of arbitrary order, too, as long as the indexing class $\mathcal{F}$ fulfilles certain restrictions

Decoupling inequality by de la Peña (1992): Let $\{X_i\}_{1 \leq i \leq n}$ be a sequence of i.i.d random variables on $\mathbb{R}$, and let $\{X_i'\}_{1 \leq i \leq n}$ be an i.i.d. copy of the sequence $\{X_i\}_{1 \leq i \leq n}$. Let f be a bounded measurable function, $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$, such that $E[f(X_i, X_j)] < \infty$. And let $\Phi : [0, \infty) \longrightarrow [0, \infty)$ be a convex non-decreasing function with $E\Phi(f(X_i, X_j)) < \infty$. Then

$$E\,\Phi\left(\Big|\sum_{i \neq j} f(X_i, X_j)\Big|\right) \leq E\,\Phi\left(8\Big|\sum_{i \neq j} f(X_i, X_j')\Big|\right)$$

If moreover f is symmetric, i.e. $f(x_1, x_2) = f(x_2, x_1)$ for all values of x_1, x_2, then also a reverse inequality holds.

$$E\,\Phi\left(\frac{1}{4}\Big|\sum_{i \neq j} f(X_i, X_j')\Big|\right) \leq E\,\Phi\left(\Big|\sum_{i \neq j} f(X_i, X_j)\Big|\right)$$

The variables at different coordinates of the domain of f in the decoupled statistic come from different independent sequences and therefore a decoupled U-statistic can be treated, conditionally on one sequence, as a sum of independent variables. In this way, the analysis of U-statistics can proceed more or less by analogy with that of sums of i.i.d. random variables.

Proof Let $\{\varepsilon_i\}_{1\le i\le n}$ be independent Rademacher variables, i.e. $P(\varepsilon_i = 1) = P(\varepsilon_i = -1) = 1/2$, independent of $\{X_i, X_i'\}_{1\le i\le n}$, and let Z_i and Z_i', $i = 1,\dots,n$, be defined as follows:

$$Z_i = \begin{cases} X_i, & \text{if } \varepsilon_i = 1 \\ X_i', & \text{if } \varepsilon_i = -1 \end{cases}, \qquad Z_i' = \begin{cases} X_i', & \text{if } \varepsilon_i = 1 \\ X_i, & \text{if } \varepsilon_i = -1 \end{cases}$$

Since $X_1,\dots,X_n$ and $X_1',\dots,X_n'$ are i.i.d., each a random variable with law P, and ε_i are independent of X_j, X_j' for all i, j, the vector $(Z_1,\dots,Z_n,Z_1',\dots,Z_n')$ is just a permutation of $(X_1,\dots,X_n,X_1',\dots,X_n')$ and

$$\begin{aligned} \mathcal{L}(Z_1,\dots,Z_n,Z_1',\dots,Z_n') &= \mathcal{L}(X_1,\dots,X_n,X_1',\dots,X_n') = P^{2n} \\ \mathcal{L}(Z_1,\dots,Z_n) &= \mathcal{L}(X_1,\dots,X_n) = P^n \end{aligned}$$

and for P^n-integrable functions f we have

$$E\, f(Z_1,\dots,Z_n) = E\, f(X_1,\dots,X_n)$$

Note also that, if $\mathcal{Z}$ is the σ-algebra generated by the X and X' variables, $\mathcal{Z} = \sigma(X_1,\dots,X_n, X_1',\dots,X_n')$, then conditional integration with respect to $\mathcal{Z}$ of any function f of the Z or Z' variables is simply integration with respect to ε and ε'. In particular, for all $i \ne j$

$$\begin{aligned} E[f(Z_i,Z_j)|\mathcal{Z}] &= E[f(Z_i,Z_j')|\mathcal{Z}] = E[f(Z_i',Z_j)|\mathcal{Z}] = E[f(Z_i',Z_j')|\mathcal{Z}] \\ &= \frac{1}{4}\Big(f(X_i,X_j) + f(X_i,X_j') + f(X_i',X_j) + f(X_i',X_j')\Big) \end{aligned}$$

These observations together with the convexity and monotony of Φ, integrability of the functions involved and Jensen's inequality, justify the following strings of inequalities which prove the theorem. For f symmetric in its entries

$$\begin{aligned} &E\,\Phi\left(\Big|\sum_{i\ne j} f(X_i,X_j')\Big|\right) \\ =\;& E\,\Phi\left(\frac{1}{2}\Big|\sum_{i\ne j} f(X_i,X_j') + \sum_{i\ne j} f(X_i',X_j)\Big|\right) \\ \le\;& \frac{1}{2}\,E\,\Phi\left(\Big|\sum_{i\ne j}\big[f(X_i,X_j') + f(X_i',X_j) + f(X_i,X_j) + f(X_i',X_j')\big]\Big|\right) \\ &+ \frac{1}{4}\,E\,\Phi\left(2\Big|\sum_{i\ne j} f(X_i,X_j)\Big|\right) + \frac{1}{4}\,E\,\Phi\left(2\Big|\sum_{i\ne j} f(X_i',X_j')\Big|\right) \\ =\;& \frac{1}{2}\,E\,\Phi\left(4\Big|\sum_{i\ne j} E[f(Z_i,Z_j)|\mathcal{Z}]\Big|\right) + \frac{1}{2}\,E\,\Phi\left(2\Big|\sum_{i\ne j} f(X_i,X_j)\Big|\right) \\ \le\;& \frac{1}{2}\,E\,\Phi\left(4\Big|\sum_{i\ne j} f(Z_i,Z_j)\Big|\right) + \frac{1}{2}\,E\,\Phi\left(2\Big|\sum_{i\ne j} f(X_i,X_j)\Big|\right) \\ =\;& \frac{1}{2}\,E\,\Phi\left(4\Big|\sum_{i\ne j} f(X_i,X_j)\Big|\right) + \frac{1}{2}\,E\,\Phi\left(2\Big|\sum_{i\ne j} f(X_i,X_j)\Big|\right) \end{aligned}$$

$$\leq \; E\,\Phi\left(4\Big|\sum_{i\neq j} f(X_i,X_j)\Big|\right)$$

proving the second part of the theorem. Note that symmetry is essential for the first identity. For f not necessarily symmetric, letting $\mathcal{X}=\sigma(X_1,\ldots,X_n)$, we have

$$\begin{aligned}
& E\,\Phi\left(\Big|\sum_{i\neq j} f(X_i,X_j)\Big|\right)\\
\leq\; & \frac{1}{2}\,E\,\Phi\left(2\Big|\sum_{i\neq j} E[f(X_i,X_j)+f(X_i',X_j)+f(X_i,X_j')+f(X_i',X_j')|\mathcal{X}]\Big|\right)\\
& +\frac{1}{2}\,E\,\Phi\left(2\Big|\sum_{i\neq j} E[f(X_i',X_j)+f(X_i,X_j')+f(X_i',X_j')|\mathcal{X}]\Big|\right)\\
\leq\; & \frac{1}{2}\,E\,\Phi\left(2\Big|\sum_{i\neq j}[f(X_i,X_j)+f(X_i',X_j)+f(X_i,X_j')+f(X_i',X_j')]\Big|\right)\\
& +\frac{1}{6}\,E\,\Phi\left(6\Big|\sum_{i\neq j} E[f(X_i',X_j)|\mathcal{X}]\Big|\right)+\frac{1}{6}\,E\,\Phi\left(6\Big|\sum_{i\neq j} E[f(X_i,X_j')|\mathcal{X}]\Big|\right)\\
& +\frac{1}{6}\,E\,\Phi\left(6\Big|\sum_{i\neq j} E[f(X_i',X_j')|\mathcal{X}]\Big|\right)\\
\leq\; & \frac{1}{2}\,E\,\Phi\left(8\Big|\sum_{i\neq j} E[f(Z_i,Z_j')|\mathcal{Z}]\Big|\right)+\frac{1}{3}\,E\,\Phi\left(6\Big|\sum_{i\neq j} f(X_i,X_j')\Big|\right)+\frac{1}{6}\,\Phi\left(6\Big|\sum_{i\neq j} E[f(X_i',X_j')]\Big|\right)\\
\leq\; & \frac{1}{2}\,E\,\Phi\left(8\Big|\sum_{i\neq j} f(Z_i,Z_j')\Big|\right)+\frac{1}{3}\,E\,\Phi\left(6\Big|\sum_{i\neq j} f(X_i,X_j')\Big|\right)+\frac{1}{6}\,\Phi\left(6\Big|\sum_{i\neq j} E[f(X_i',X_j')]\Big|\right)\\
=\; & \frac{1}{2}\,E\,\Phi\left(8\Big|\sum_{i\neq j} f(X_i,X_j')\Big|\right)+\frac{1}{3}\,E\,\Phi\left(6\Big|\sum_{i\neq j} f(X_i,X_j')\Big|\right)+\frac{1}{6}\,\Phi\left(6\Big|\sum_{i\neq j} E[f(X_i,X_j')]\Big|\right)\\
\leq\; & E\,\Phi\left(8\Big|\sum_{i\neq j} f(X_i,X_j')\Big|\right)
\end{aligned}$$

proving the first part of the theorem. □

The most interesting aspect of decoupling is the possibility of randomizing a degenerate U-statistic. A symmetric function f is called degenerate, if $E[f(X,Y)|Y=y]=E[f(X,Y)]$ for all values y in the domain of $f(x,.)$, whereas $f(x,y)$ is not a constant function.
The Randomization theorem to be stated below can be considered as an extension of the following inequality for univariate functions, Zinn (1985), (ε_i defined as before):

$$2^{-p}E\Big|\sum_{i=1}^n \varepsilon_i f(X_i)\Big|^p \leq E\Big|\sum_{i=1}^n f(X_i)\Big|^p \leq 2^p E\Big|\sum_{i=1}^n \varepsilon_i f(X_i)\Big|^p, \quad p\geq 1$$

which is connected to the well known Symmetrization lemma (e.g. Pollard (1984), p. 14). The symbol $\simeq$ may denote a two sided inequality up to multiplicative constants that depend only on the exponent p.

Randomization theorem Let f be a degenerate function with $E[f(X,Y)] = 0$, $E|f(X,Y)|^p < \infty$, X, X', ε as before and ε' an independent copy of ε, $p \geq 1$. Then

$$\begin{aligned} E\Big|\sum_{i\neq j} f(X_i,X_j)\Big|^p &\simeq E\Big|\sum_{i\neq j} \varepsilon_i\varepsilon_j f(X_i,X_j)\Big|^p \\ &\simeq E\Big|\sum_{i\neq j} \varepsilon_i\varepsilon_j' f(X_i,X_j)\Big|^p \\ &\simeq E\Big|\sum_{i\neq j} \varepsilon_i\varepsilon_j' f(X_i,X_j')\Big|^p \end{aligned}$$

Proof Since f is degenerate and centered, we have $E[f(X,Y)|Y] = 0$. For iterated expectations, the Randomization inequality for univariate functions gives:

$$\begin{aligned} E\Big|\sum_{i\neq j} f(X_i,X_j')\Big|^p &= E\left[E\Big[\Big|\sum_{i\neq j} f(X_i,X_j')\Big|^p\Big|X_j'\Big]\right] \\ &\simeq E\left[E\Big[\Big|\sum_{i\neq j} \varepsilon_i f(X_i,X_j')\Big|^p\Big|X_j'\Big]\right] \\ &= E\Big|\sum_{i\neq j} \varepsilon_i f(X_i,X_j')\Big|^p \\ &= E\left[E\Big[\Big|\sum_{i\neq j} \varepsilon_i f(X_i,X_j')\Big|^p\Big|X_i\Big]\right] \\ &\simeq E\left[E\Big[\Big|\sum_{i\neq j} \varepsilon_i\varepsilon_j' f(X_i,X_j')\Big|^p\Big|X_i\Big]\right] \\ &= E\Big|\sum_{i\neq j} \varepsilon_i\varepsilon_j' f(X_i,X_j')\Big|^p \end{aligned}$$

The equivalences of the theorem follow by continued application of the Decoupling inequality.

□

Now we proceed to the aim of the whole section, the proof of a Bernstein type inequality for degenerate U-statistics, Proposition 2.3 (c) of Arcones, Giné (1993):

Proposition Let $\max|f| \leq c$, $E[f(X,Y)] = 0$, f symmetric and degenerate and $\sigma^2 := E[f^2(X,Y)]$. Then there are constants c_1 and c_2 not depending on f, such that for any $t > 0$:

$$P\left(\Big|\frac{1}{n}\sum_{i\neq j} f(X_i,X_j)\Big| > t\right) \leq c_1 \exp\left\{\frac{-c_2 t}{\sigma + (ct^{1/2}n^{-1/2})^{2/3}}\right\}$$

Proof As for Bernstein's inequality, this proof is just a subtle application of Markov's inequality. The difficulty lies in finding a bound to the exponential moment of the U-statistic.

Let $a_{ij} \in \mathbb{R}$, ε be defined as above and $X := \sum_{i \neq j} \varepsilon_i \varepsilon_j \, a_{ij}$, $s^2 := \sum_{i \neq j} a_{ij}^2$.
In order to bound $E[\exp\{t|X|^\alpha\}]$, where $t > 0$ and $0 < \alpha < 1$, we can write:

$$\begin{aligned} E\left[\exp\{t|X|^\alpha\}\right] &= E\left[\exp\{|\lambda X/(\alpha s)|^\alpha (\alpha s/\lambda)^\alpha t\}\right] \\ &\leq E\left[\exp\Big\{|\lambda X/s| + (1-\alpha)(\alpha s/\lambda)^{\alpha/(1-\alpha)} t^{1/(1-\alpha)}\Big\}\right] \end{aligned}$$

with any constant λ, say $\lambda > 0$, (since for real a and b, $1 < p < \infty$: $|ab| \leq |a|^p/p + (p-1)|b|^{p/(p-1)}/p$). Now we expand $\exp\{|\lambda X/s|\}$ into a series:

$$E\left[\exp\{|\lambda X/s|\}\right] = 1 + E\left[|\lambda X/s|\right] + \frac{1}{2}\, E\left[|\lambda X/s|^2\right] + \sum_{k \geq 3} \frac{1}{k!}\, E\left[|\lambda X/s|^k\right]$$

and apply kind of Rosenthal's inequality to the higher moments (Borell (1979)).

$$\text{For } \; 1 < q < p < \infty: \quad E^{1/p}|X|^p \;\leq\; \left(\frac{p-1}{q-1}\right) E^{1/q}|X|^q, \qquad \text{such that}$$

$$\begin{aligned} E\left[\exp\{|\lambda X/s|\}\right] &= 1 + E\left[|\lambda X/s|\right] + \frac{1}{2}\, E\left[|\lambda X/s|^2\right] + \sum_{k \geq 3} \frac{(k-1)^k}{k!}\, E^{k/2}\left[|\lambda X/s|^2\right] \\ &= 1 + \lambda + \frac{\lambda^2}{2} + \sum_{k \geq 3} \frac{(k-1)^k \lambda^k}{k!} \end{aligned}$$

which is a polynomial in λ, and finite when $\lambda < e^{-1}$. Together with the previous inequality we have:

$$\begin{aligned} E\left[\exp\{t|X|^\alpha\}\right] &\leq P(\lambda)\; \exp\left\{(1-\alpha)(\alpha s/\lambda)^{\alpha/(1-\alpha)} t^{1/(1-\alpha)}\right\} \\ &=: c_1 \exp\left\{c_2\, (s^\alpha t)^{1/(1-\alpha)}\right\} \end{aligned}$$

Take $\alpha = 2/3$ and $a_{ij} = \frac{1}{n} f(X_i, X_j)$, such that the exponent of s equals 2 and $X = \frac{1}{n} \sum_{i \neq j} f(X_i, X_j)$. Let $\Phi(.) = \exp\{t|.|^\alpha\}$, a convex, non-decreasing function, and apply the Decoupling inequality and the Randomization theorem.

$$\begin{aligned} E\left[\exp\Big\{t\Big|\frac{1}{n}\sum_{i \neq j} f(X_i, X_j)\Big|^{2/3}\Big\}\right] &\leq c_1' E\left[\exp\Big\{t\Big|\frac{1}{n}\sum_{i \neq j} \varepsilon_i \varepsilon_j\; f(X_i, X_j)\Big|^{2/3}\Big\}\right] \\ &\leq c_1 c_1' E\left[\exp\Big\{c_2\, t^3 \frac{1}{n^2} \sum_{i \neq j} f^2(X_i, X_j)\Big\}\right] \\ &= c_1'' \exp\Big\{c_2\, t^3\; \frac{n-1}{n}\; \sigma^2\Big\} E\left[\exp\Big\{c_2\, t^3 \frac{1}{n^2} \sum_{i \neq j} (f^2(X_i, X_j) - \sigma^2)\Big\}\right] \\ &\leq c_1'' \exp\Big\{c_2'\, t^3\, \sigma^2\Big\} E\left[\exp\Big\{c_2\, t^3 \frac{1}{n^2} \sum_{i \neq j} (f^2(X_i, X_j) - \sigma^2)\Big\}\right] \end{aligned}$$

We take advantage of the fact that $\frac{2}{n(n-1)}\sum_{i\neq j} f(X_i,X_j) = \frac{1}{n!}\sum_{\mathcal{P}}\frac{1}{\lfloor n/2\rfloor}\sum_{l=1}^{\lfloor n/2\rfloor} f(X_{i_{2l-1}},X_{i_{2l}})$, where $\mathcal{P}=\{(i_1,\ldots,i_n)|1\le i_j\le n, i_j\neq i_k \text{ for } j\neq k\}$ is the set of permutations of $(1,\ldots,n)$. In this case, an inequality is valid, that is given for example in Höffding (1963) and in Serfling (1980):

$$E\left[\exp\Big\{c_2\,t^3\frac{1}{n^2}\sum_{i\neq j}(f^2(X_i,X_j)-\sigma^2)\Big\}\right] \;\le\; E\left[\exp\Big\{4\,c_2\,t^3\frac{1}{n}\sum_{i=1}^{\lfloor n/2\rfloor}(f^2(X_{2i-1},X_{2i})-\sigma^2)\Big\}\right]$$

The random variables in the right-hand side's sum are now i.i.d. and we make use of some form of Bernstein's inequality for $\xi_1,\ldots,\xi_m$ i.i.d. with $|\xi|\le d$, $E\xi=0$ and $E\xi^2=\tau^2$:

$$E\left[\exp\Big\{v\,\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\xi_i\Big\}\right]\le\exp\Big\{\frac{v^2\tau^2}{2-2/3\;v\,d\,m^{-1/2}}\Big\},\qquad\text{for }\;|v|<\frac{3\sqrt{m}}{d}$$

Set $\xi:=f^2(X_{2i-1},X_{2i})-\sigma^2$, $m:=\lfloor n/2\rfloor$, $v:=4c_2t^3\lfloor n/2\rfloor^{1/2}n^{-1}$, $d:=c^2$ and, $\tau^2:=c^2\sigma^2$. Then

$$E\left[\exp\Big\{4\,c_2\,t^3\frac{1}{n}\sum_{i=1}^{\lfloor n/2\rfloor}(f^2(X_{2i-1},X_{2i})-\sigma^2)\Big\}\right]\;\le\;\exp\left\{\frac{c_2^2\,t^6c^2\sigma^2}{n/4-c_2t^3c^2/3}\right\}$$

When we combine all inequalities developed so far, we obtain a bound for the first exponential moment of our U-statistic:

$$E\left[\exp\Big\{t\Big|\frac{1}{n}\sum_{i\neq j}f(X_i,X_j)\Big|^{2/3}\Big\}\right]\;\le\;c_1''\exp\Big\{c_2'\,t^3\,\sigma^2\Big\}\exp\left\{\frac{c_2^2\,t^6c^2\sigma^2}{n/4-c_2t^3c^2/3}\right\}$$

Finally, we employ Markov's inequality:

$$\begin{aligned}
P\left(\Big|\frac{1}{n}\sum_{i\neq j}f(X_i,X_j)\Big|>u\right) &= P\left(\exp\Big\{t\Big|\frac{1}{n}\sum_{i\neq j}f(X_i,X_j)\Big|^{2/3}\Big\}>\exp\{t\,u^{2/3}\}\right)\\
&\le \exp\{-tu^{2/3}\}E\left[\exp\Big\{4\,c_2\,t^3\frac{1}{n}\sum_{i=1}^{\lfloor n/2\rfloor}(f^2(X_{2i-1},X_{2i})-\sigma^2)\Big\}\right]\\
&\le c_1''\exp\left\{-tu^{2/3}+c_2't^3\sigma^2+\frac{c_2^2\,t^6c^2\sigma^2}{n/4-c_2t^3c^2/3}\right\}
\end{aligned}$$

The value $t^*:=u^{1/3}(3c_2')^{-1/2}\sigma^{-1}$ minimizes the sum of the first two terms in the exponent of the bound, giving $-t^*u^{2/3}+c_2't^{*3}\sigma^2=-c_3\,u\sigma^{-1}$. This is the leading term, whenever the third summand is smaller than a small constant C times $c_3\,u\sigma^{-1}\iff c_3'uc^2(n-uc^2)^{-1}>C$, with appropriate c_3'. If this is not fulfilled, the exponent is of order $-c_2''(nu^2/c^2)^{1/3}$. Thereby the proposition is proven. □

We apply the Bernstein type inequality in Part I, Lemma I.4.2 to the degenerate U-statistics $U_{\varphi_{st}}(X_i,X_j)=\varphi_{st}(X_i-X_j)-E[\varphi_{st}(X_i-X_j)|X_i]-E[\varphi_{st}(X_i-X_j)|X_j]+E[\varphi_{st}(X_i-X_j)]$ and $U_{\psi_{st}}(X_i,X_j)=\psi_{st}(X_i-X_j)-E[\psi_{st}(X_i-X_j)|X_i]-E[\psi_{st}(X_i-X_j)|X_j]+E[\psi_{st}(X_i-X_j)]$, respectively. The maxima of $U_{\varphi_{st}}$ and $U_{\psi_{st}}$ are obviously bounded by 4 times the maximum of the wavelet functions. For $E|\varphi_{st}|^2$ we plug in $\|f\|_\infty \|\varphi_{st}\|_2^2 = \frac{1}{2\pi}\|f\|_\infty$ (recall that in Part I, f is the density to be estimated), the same for ψ_{st}. A better bound could have been found through $E|\varphi_{st}|^2 \leq \frac{1}{2\pi}\|f\|_2^2$, allowing to drop the assumption $\|f\|_\infty < \infty$ in Lemma I.4.2. Nevertheless, such a bound is not attainable for the wavelet functions φ'_{st} and ψ'_{st} in Lemma I.4.4. And therefore the gain in Lemma I.4.2 is of no use for Theorem I.2.1.

A.3 Exponential inequality for quadratic forms

An extensive amount of work investigates limit theorems of large deviations of sums of independent variables, which is quite natural, because the simplest case, to which powerful analytical methods are applicable, enables the beholder to get a clear and complete picture of the behavior of probabilities of large deviations. Different assumptions on the distribution of the summands have been considered: characteristic functions which are analytical in a neighborhood of 0 (Cramér condition); existence of all moments without analyticity (Linnik case); the case of moderate deviations, when only a finite number of moments exists; summands belonging to the domain of attraction of the stable law with $\alpha < 2$; conditions for super-large deviations.
The ideas that work for sums of independent variables can also be applied to sums of dependent variables, polynomial forms and multiple stochastic integrals of random processes. Saulis and Statulevičius (1991) address applications of the method of cumulants to limit theorems of large deviations, which proves its usefulness at the highest stage.
We will give an account of the exponential inequality for polynomial forms, Lemma 2.1, Bentkus, Rudzkis (1980), cited in Saulis, Statulevičius (1991) as Lemma 2.4, p. 19, of which the exponential inequality for quadratic forms, used in Lemma II.2.2, Part II of the present work, is a special case.

Let the k^{th} cumulant of a random variable X be denoted by $\Gamma_k(X)$, i.e.

$$\Gamma_k(X) := i^{-k} \frac{d^k}{d\omega^k} \ln E e^{iX\omega}|_{\omega=0}\,, \qquad k \in \mathbb{N}$$

It is often less involved and more effective to yield result by means of cumulants than by means of moments. Because the k^{th} cumulant of a sum of independent random variables is just the sum of the k^{th} cumulants, while in contrast the k^{th} moment does not admit such a simple representation. Furthermore note, that none but the first cumulant changes, when adding a constant to a random variable.
In Lemma II.2.2, Part II, we are looking upon quadratic forms $\varepsilon^T A \varepsilon$ of i.i.d random variables $\varepsilon_i \sim \mathcal{N}(0,1)$. Let λ_i be the eigenvalues of A. Then by spectral decomposition we have $\varepsilon^T A \varepsilon = \sum \lambda_i \zeta_i$ a weighted sum of independent χ^2-distributed random variables. The cumulants of the χ^2-distribution are for example given in Johnson, Kotz (1970), p. 168. We derive $\Gamma_k(\varepsilon^T A \varepsilon) = \sum \lambda_i^k \, \Gamma_k(\zeta_i) = \sum \lambda_i^k \, 2^{k-1}(k-1)!$

$$\begin{aligned}
|\Gamma_k(\varepsilon^T A \varepsilon)| &\leq 2^{k-1}(k-1)! \sum_{i=1}^{n} \lambda_i^2 \max_j |\lambda_j|^{k-2} \\
&= \frac{k!}{2}\, 2 \sum_{i=1}^{n} \lambda_i^2 \, \frac{2}{k} \left(2 \max_j |\lambda_j|\right)^{k-2} \\
&\leq \frac{k!}{2} \, \frac{H}{\Delta^{k-2}}
\end{aligned}$$

where $H := 2\sum\lambda_i^2 = 2\,\mathrm{tr}A^TA$ and $\Delta^{-1} := 2\max|\lambda_j| \le 2(\max_j\sum_i|A_{ij}|\cdot\max_i\sum_j|A_{ij}|)^{1/2}$ (Lütkepohl (1996), p.112). Disposing of this approximation, we can apply:

Lemma for quadratic forms (Bentkus, Rudzkis (1980) Lemma 2.1) Let ξ be an arbitrary random variable with $E\xi = 0$. Denote the cumulants $\Gamma_k(\xi)$ of ξ by γ_k. If there exist $\gamma \ge 0$, $H > 0$ and $\Delta > 0$ such that

$$|\gamma_k| \le \left(\frac{k!}{2}\right)^{1+\gamma}\frac{H}{\Delta^{k-2}} \qquad \text{for } \ k \ge 2$$

then for all $x \ge 0$:

$$P(|\xi| \ge x) \le 2\exp\left\{-\frac{x^2}{2H + 2x^{\frac{1+2\gamma}{1+\gamma}}\Delta^{-\frac{1}{1+\gamma}}}\right\}$$

For $\xi = \varepsilon^TA\varepsilon - E\varepsilon^TA\varepsilon$ we even have $\gamma = 0$. And with A symmetric, $\mathrm{tr}\,A^TA = \mathrm{tr}\,A^2$ and $\max|\lambda_j| \le \max_j\sum_i|A_{ij}| = \|A\|_\infty$. Adjusting the form of Bentkus/Rudzkis' inequality to Lemma II.2.2, we result in

$$P(|\varepsilon^TA\varepsilon - E\varepsilon^TA\varepsilon| \ge x) \le 2\exp\left\{-\frac{x^2/2}{2\mathrm{tr}\,A^2 + 2x\|A\|_\infty}\right\}$$

This lemma cannot only be formulated so as to apply to quadratic but also to arbitrary polynomial forms. In the special case of a linear form $a^T\varepsilon$ in i.i.d. random variables with $\varepsilon_i \sim \mathcal{N}(0,1)$ (for which we use it in Lemma II.2.4) the cumulants are especially easy to calculate, since $\Gamma_1(\varepsilon_i) = 0$, $\Gamma_2(\varepsilon_i) = 1$ and and again $\Gamma_k(\varepsilon_i) = 0$ for $k \ge 3$. Bentkus/Rudzkis' lemma gives exactly the sharp version of Bernstein's inequality (Petrov (1987)) that is known to hold for i.i.d. standard-normal random variables:

$$P(|a^T\varepsilon| \ge x) \le 2\exp\left\{-\frac{x^2}{2\|a\|_2^2}\right\}$$

Several preliminary remarks will contribute to the clarity of the proof of the lemma, for reference see Bentkus, Rudzkis (1980) or Saulis, Statulevičius (1991):

1. For a non-negative integer m and a function $g : \mathbb{R} \longrightarrow \mathbb{R}$ of x, having m derivatives at $x = 0$, define

$$g_m(x) := \sum_{k=0}^m \frac{1}{k!}\,g^{(k)}(0)\,x^k\,, \qquad x \in \mathbb{R}$$

which are the first m members of a Taylor expansion. In particular,

$$\exp_m\{x\} := \sum_{k=0}^m \frac{1}{k!}\,x^k\,, \qquad x \in \mathbb{R}$$

Note that the function $\exp_{2n}$ is positive for every $n \in \mathbb{N}$. For $x \ge 0$, this is obvious. Otherwise define $a_k := \frac{x^{2k-1}}{(2k-1)!} + \frac{x^{2k}}{(2k)!}$. When $x \le -2n$, then $a_k \ge 0$ for $1 \le k \le n$ and $\exp_{2n}\{x\} = 1 + a_1 + \ldots + a_n > 0$. When $-2n < x < 0$, then $a_k < 0$ for $k \ge n-1$ and $e^x = \exp_{2n}\{x\} + a_{n+1} + a_{n+2} + \ldots$, such that $\exp_{2n}\{x\} > 0$.

2. The inequality

$$\frac{\exp_n\{x\}}{\exp_{2n}\{2x\}} \le \exp\{-x\}$$

is valid for $0 \le x \le 0.6$ if $n = 1$, for $0 \le x \le 1.4$ if $n = 2$, and for $0 \le x \le 0.8n$ if $n \ge 3$.

3. Over the interval $(2.5, 6.4]$, the value of the function

$$f(t) = \frac{1}{t} - \frac{1}{e^{t/2} - 1}, \qquad t > 0$$

is less than 0.172.

Proof of Bentkus/Rudzkis' lemma Since $E\xi^{2n} < \infty$ for all $n \in \mathbb{N}$, and since $\exp_{2n}\{x\}$ is non-negative on $\mathbb{R}$ and monotonously increasing on $\mathbb{R}^+$, we can apply Markov's inequality to yield for all $h \ge 0$ and all $x \ge 0$:

$$P(|\xi| \ge x) \le 2P\Big(\exp_{2n}\{h\xi\} \ge \exp_{2n}\{hx\}\Big) \le 2\exp_{2n}^{-1}\{hx\} \sum_{k=0}^{2n} \frac{1}{k!}\, E\xi^k h^k$$

Denote

$$g(x) := \exp_n\left\{ \sum_{r=2}^{2n} \frac{1}{r!}\, \gamma_r x^r \right\}, \qquad x \in \mathbb{R}$$

According to the calculation rules for cumulants, for any $k \in \mathbb{N}^+$ it holds that

$$E\xi^k = \sum_{q=1}^{k} \sum_{r_1+\ldots+r_p=k} \frac{1}{q!} \frac{k!}{r_1! \cdots r_q!}\, \gamma_{r_1} \cdots \gamma_{r_q}$$

Under condition $E\xi = 0$, this gives for any $k = 2, 3, \ldots$

$$\begin{aligned} E\xi^k &= \sum_{q=1}^{k} \sum_{r_1+\ldots+r_p=k, r_j \ge 2} \frac{1}{q!} \frac{k!}{r_1! \cdots r_q!}\, \gamma_{r_1} \cdots \gamma_{r_q} \\ &= \sum_{q=1}^{k} \frac{1}{q!} \frac{d^k}{dx^k} \left(\sum_{r=2}^{2n} \frac{1}{r!}\, \gamma_r x^r \right)^q \Bigg|_{x=0} \\ &= \frac{d^k}{dx^k}\, g(x)\Big|_{x=0} \end{aligned}$$

Consequently,

$$\sum_{k=0}^{2n} \frac{1}{k!}\, E\xi^k h^k = g_{2n}(h) \le \exp_n \left\{ \sum_{r=2}^{2n} \frac{1}{r!}\, |\gamma_r| h^r \right\}$$

by virtue of g analytical and $g^{(k)}(x) > 0$ on $\mathbb{R}^+$ for all k. We substitute for g_{2n}, such that for all $h \ge 0$:

$$P(|\xi| \ge x) \le 2\exp_{2n}^{-1}\{hx\}\, g_{2n}(h) \le 2\exp_{2n}^{-1}\{hx\} \exp_n \left\{ \sum_{k=2}^{2n} \frac{1}{k!}\, |\gamma_k| h^k \right\}, \quad x \ge 0$$

Applying the assumption $|\gamma_k| \le (\frac{k!}{2})^{1+\gamma} \frac{H}{\Delta^{k-2}}$, it suffices to consider the case $H = 1$. Otherwise we could introduce the random variable $\widetilde{\xi} = \xi/H$, for which the assumption would be fulfilled with $\widetilde{H} = 1$ and $\widetilde{\Delta} = \Delta\sqrt{H}$. The estimate for $P(|\xi| \ge x)$ could be obtained from the one for $P(|\widetilde{\xi}| \ge x\sqrt{H})$.

Passing forward by inserting the assumption into the bound of $P(|\xi| \geq x)$, for $n \in \mathbb{N}^+$ and all $h \geq 0$ we arrive at

$$P(|\xi| \geq x) \ \leq \ \frac{2\exp_n\left\{\sum\limits_{k=2}^{2n} \frac{1}{k!}\,|\gamma_k| h^k\right\}}{\exp_{2n}\{hx\}} \ \leq \ \frac{2\exp_n\left\{\frac{1}{2}\,h^2 \sum\limits_{k=2}^{2n} \left(\frac{h}{\Delta}\right)^{k-2} \left(\frac{k!}{2}\right)^{\gamma}\right\}}{\exp_{2n}\{hx\}} \tag{$*$}$$

On the other hand, also applying Markov's inequality and taking into account that $E\xi = 0$ and $E\xi^2 \leq H = 1$, we have for all $x \geq 0$

$$P(|\xi| \geq x) \leq P\Big((1+x\xi)^2 \geq (1+x^2)^2\Big) \leq \frac{1 + 2xE\xi + x^2 E\xi^2}{(1+x^2)^2} \leq \frac{1}{1+x^2} \tag{$**$}$$

Choose

$$h = \frac{x(x\Delta)^{1/(1+\gamma)}}{x^2 + (x\Delta)^{1/(1+\gamma)}}$$

and n such that the condition $0.8(n-1) < hx/2 \leq 0.8n$ is fulfilled. First consider the case $hx > 6.4$, i.e. $n \geq 5$. Then for $k = 2, 3, \ldots, 2n$

$$\frac{k!}{2} \leq (hx)^{k-2}$$

(In fact, since $hx > 1.6(n-1)$, this follows from the easily verifiable inequality $(2n)!/2 \leq (1.6(n-1))^{2n-2}$, $n = 5, 6, \ldots$)
By substituting the chosen values of h and n, we get

$$\frac{1}{2}\,h^2 \sum_{k=2}^{2n} \left(\frac{h}{\Delta}\right)^{k-2} \left(\frac{k!}{2}\right)^{\gamma} \leq \frac{1}{2}\,h^2 \sum_{k=2}^{2n} \left(\frac{h}{\Delta}(hx)^{\gamma}\right)^{k-2} \leq \frac{h^2}{2}\,\frac{1}{1-q} = \frac{hx}{2}$$

since

$$q = \left(\frac{h}{\Delta}(hx)^{\gamma}\right)^{1/(1+\gamma)} = \frac{x^2}{x^2 + (x\Delta)^{1/(1+\gamma)}} < 1$$

In connection with $(*)$, and using remark 2, we obtain

$$P(|\xi| \geq x) \leq 2\exp_{2n}^{-1}\{hx\}\exp_n\left\{\frac{hx}{2}\right\} \leq 2\exp\left\{-\frac{hx}{2}\right\}$$

in other words the claim of the lemma.
Now consider the case $0 < h \leq 6.4$. We split this into two subcases: 1) when condition

$$\frac{1}{1+x^2} \leq \exp\left\{-\frac{hx}{2}\right\}$$

is fulfilled and 2) when it is not fulfilled. In the first subcase the claim follows from $(**)$. In the second subcase, since $x > h$

$$\exp\left\{\frac{hx}{2}\right\} > 1 + x^2 > 1 + hx$$

which implies that $hx > 2.5$. Thus, there remains to consider $2.5 < hx \leq 6.4$ under $\exp\{hx/2\} > 1 + hx$. In this case, the claim is obtained from $(*)$ by setting $n = 4$. By virtue of $\exp\{hx/2\} > 1 + hx$ we have $x^2 < \exp\{hx/2\} - 1$, therefore with remark 3

$$\frac{h}{\Delta} = q \cdot \left(\frac{1}{hx} - \frac{1}{x^2}\right)^{\gamma} < q \cdot f^{\gamma}(hx) \leq q \cdot 0.172^{\gamma}$$

and

$$\frac{1}{2}h^2 \sum_{k=2}^{8} \left(\frac{h}{\Delta}\right)^{k-2} \left(\frac{k!}{2}\right)^{\gamma} \leq \frac{1}{2}h^2 \sum_{k=2}^{8} q^{k-2} \left(\frac{k!}{2} \cdot 0.172^{k-2}\right)^{\gamma} \leq \frac{h^2}{2(1-q)} = \frac{hx}{2}$$

since $(k!/2) \cdot 0.172^{k-2} < 1$ for all $k = 2, 3, \ldots, 8$. Again this yields the claim of the lemma. □

A.4 Van Trees inequality

The van Trees inequality (van Trees (1968)) is a Bayesian version of the well known Cramér-Rao bound. This inequality is used by Gill and Levit (1995) to prove that the limiting distribution of any regular estimator cannot have a variance less than the classical information bound, under minimal regularity conditions. They show how minimax convergence rates can be derived in various non- and semi-parametric problems from the van Trees inequality. Multivariate versions of the inequality are developed and applications given. Especially interesting to us is the example on the empirical distribution function, which gave the impetus to Golubev, Levit (1996) and Schipper (1996) to work out a new and elementary proof of the asymptotic minimax bounds of the mean square error of an arbitrary estimator of a distribution function and of a density function, repsectively, already obtained earlier by the help of local asymptotic normality.

First we prove the van Trees inequality in its simplest version (Gill and Levit (1995)), later on we also give a generalization and multivariate extensions and an explanation of how the van Trees inequality is applied to our problem.

Let $(\mathcal{X}, \mathcal{F}, P_\theta | \theta \in \Theta)$ be a dominated family of distributions on some sample space $\mathcal{X}$; denote the dominating measure by μ. Take the parameter space Θ to be a closed interval on the real line, $\Theta = [\theta_0, \theta_1]$. Let $f(x|\theta)$ denote the density of P_θ with respect to μ. Let Λ be some probability distribution on Θ with a density $\lambda(\theta)$ with respect to Lebesgue measure. Suppose that λ and $f(x|\cdot)$ are both absolutely continuous (μ-almost surely), and that λ converges to 0 at the endpoints of the interval Θ.

Let $\widetilde{\theta} = \widetilde{\theta}(X)$ denote any estimator of θ, $X \sim P_\theta$. Write E_{f_θ} for expectation with respect to f_θ. When $\boldsymbol{\theta}$ is drawn from the distribution Λ, and conditional on $\boldsymbol{\theta} = \theta$, X from P_θ, write $E_\lambda E_{f_\theta}$ for expectation with respect to the ensuing distribution of X and $\boldsymbol{\theta}$.

Apart from the absolute continuity of f as function of θ, the last assumption is the usual

$$E_{f_\theta}\left[\frac{\partial}{\partial\theta} \ln f(X|\theta)\right] = 0$$

Define further

$$\begin{aligned} I_{f_\theta} &:= E_{f_\theta}\left[\left(\frac{\partial}{\partial\theta} \ln f(X|\theta)\right)^2\right] \\ I_\lambda &:= E_\lambda E_{f_\theta}\left[\left(\frac{d}{d\theta}\,\lambda(\theta)\right)^2\right] \end{aligned}$$

the Fisher information for θ and for a location parameter in Λ, respectively. Now

$$\int \frac{\partial}{\partial\theta}\Big(f(x|\theta)\lambda(\theta)\Big)\,d\theta = f(x|\theta)\lambda(\theta)\Big|_{\theta_0}^{\theta_1} = 0$$

by convergence of λ to 0 at the endpoints of Θ, and by partial integration and the same fact again

$$\int \theta \, \frac{\partial}{\partial\theta}\Big(f(x|\theta)\lambda(\theta)\Big)d\theta = \theta f(x|\theta)\lambda(\theta)\Big|_{\theta_0}^{\theta_1} - \int f(x|\theta)\lambda(\theta)d\theta = -\int f(x|\theta)\lambda(\theta)d\theta$$

Using both these equalities,

$$\iint \Big(\widehat{\theta}(x) - \theta\Big)\,\frac{\partial}{\partial\theta}\Big(f(x|\theta)\lambda(\theta)\Big)d\theta\,\mu(dx) = \iint f(x|\theta)\lambda(\theta)d\theta\,\mu(dx) = 1$$

Cauchy-Schwarz now gives

$$\iint \Big(\widehat{\theta}(x) - \theta\Big)^2 f(x|\theta)\lambda(\theta)d\theta\,\mu(dx) \iint \Big(\frac{\partial}{\partial\theta}\ln\Big(f(x|\theta)\lambda(\theta)\Big)\Big)^2 f(x|\theta)\lambda(\theta)d\theta\,\mu(dx) \geq 1$$

By assumption $E_{f_\theta}\left[\frac{\partial}{\partial\theta}\ln f(X|\theta)\right] = 0$, the second factor in the left-hand side's product reduces to

$$\iint \frac{1}{f(x|\theta)\lambda(\theta)}\Big(\frac{\partial}{\partial\theta}f(x|\theta)\cdot\lambda(\theta) + f(x|\theta)\cdot\frac{d}{d\theta}\lambda(\theta)\Big)^2 d\theta\,\mu(dx) = \int I_{f_\theta}\lambda(\theta)d\theta + 0 + I_\lambda$$

Dividing out and abbreviating the notation gives the final inequality

$$E_\lambda E_{f_\theta}\left[\widetilde{\theta}(X) - \theta\right]^2 \geq \frac{1}{E_\lambda[I_{f_\theta}] + I_\lambda}$$

When applying this inequality to an estimator based on n i.i.d. random variables, the information for θ is n times the information for one observation.
A more general inequality for estimating an absolutely continuous function ψ of θ is easily obtained in exactly the same way. Replacing θ by $\psi(\theta)$,

$$\begin{aligned}\int \psi(\theta)\,\frac{\partial}{\partial\theta}\Big(f(x|\theta)\lambda(\theta)\Big)d\theta &= \psi(\theta)f(x|\theta)\lambda(\theta)\Big|_{\theta_0}^{\theta_1} - \int \frac{d}{d\theta}\psi(\theta)\; f(x|\theta)\lambda(\theta)d\theta \\ &= -\int \frac{d}{d\theta}\psi(\theta)\; f(x|\theta)\lambda(\theta)d\theta\end{aligned}$$

Replacing $\widehat{\theta}(x)$ by $\widehat{\psi}(x)$ in the subsequent development gives

$$E_\lambda E_{f_\theta}\left[\widetilde{\psi}(X) - \theta\right]^2 \geq \frac{E_\lambda^2\left[\frac{d}{d\theta}\psi(\theta)\right]}{E_\lambda[I_{f_\theta}] + I_\lambda}$$

See Gill, Levit (1995) for examples and an asymptotic Cramér-Rao bound.
Next we give a very general version of the multivariate van Trees inequality, allowing for an estimator of a p-dimensional vector-valued function ψ of an s-dimensional vector parameter θ, and furthermore involving arbitrary choices of certain matrix weight functions. Suppose from the start that we have n i.i.d. observations X_i from a common distribution P_θ with density $f(x, \theta)$ with respect to some measure μ (all on an arbitrary measure space $\mathcal{X}$). Suppose $\theta \in \Theta \subseteq \mathbb{R}^s$.
Next choose a prior density (with repsect to Lebesgue measure) $\lambda(\theta)$, a symmetric $p \times p$ matrix function $B(\theta)$, and a $p \times s$ matrix function $C(\theta)$. A number of regularity conditions on f, λ, B and C are necessary.

In the sequel, a real measurable function $g(\theta)$, $\theta \in \Theta$, is said to be *nice* if, for each j, it is absolutely continuous in θ_j for almost all values of the other components of θ. θ and ψ are treated as column vectors, partial derivatives with respect to the components of θ are set out in rows, such that $\partial\psi/\partial\theta$ is a $p \times s$ matrix.

Assumptions

I. f is measurable in x, θ and *nice* in θ for almost all x.

II. The Fisher information matrix $I_{f_\theta} = E_{f_\theta}\left[\left(\frac{\partial \ln f(x,\theta)}{\partial\theta}\right)^T\left(\frac{\partial \ln f(x,\theta)}{\partial\theta}\right)\right]$ exists and the diagonal of I_{f_θ} is locally integrable in θ.

III. The components of ψ and C are *nice*.

IV. B is positive definite; suppose $B(\theta) = A(\theta)^T A(\theta)$ for a $p \times p$ matrix $A(\theta)$.

V. λ is *nice*; Θ is compact with boundary which is piecewise C_1-smooth; λ is positive on the interior of Θ and 0 on its boundary.

Assumptions I and II imply (Borovkov (1984); Borovkov and Sakhanenko (1980)) that the expected score vector is 0 for almost all θ: $E_{f_\theta}\left[\frac{\partial \ln f(x,\theta)}{\partial\theta}\right] = 0$.

Multivariate van Trees inequality Under assumptions I-V, for any estimator $\widetilde{\psi}_n$ of $\psi(\theta)$

$$E_\lambda E_{f_\theta}\left[\left(\widetilde{\psi}_n - \psi(\theta)\right)^T B(\theta)^{-1}\left(\psi_n - \psi(\theta)\right)\right] \geq \frac{E_\lambda^2\left[\text{tr } C(\theta)\left(\frac{\partial\psi}{\partial\theta}\right)^T\right]}{nE_\lambda\left[\text{tr } B(\theta)^T C(\theta) I_{f_\theta} C(\theta)^T\right] + I'_\lambda}$$

where

$$I'_\lambda = E_\lambda\left[\frac{1}{\lambda^2(\theta)}\sum_{i,j,k,l} B_{ij}(\theta)\frac{\partial}{\partial\theta_k}\left(C_{ik}(\theta)\lambda(\theta)\right)\frac{\partial}{\partial\theta_l}\left(C_{il}(\theta)\lambda(\theta)\right)\right]$$

This inequality is applied in Part III, proof of Theorem III.3.3 to the problem of finding a lower bound to $E_\lambda E_{f_\theta}[\text{Re}\widetilde{\widehat{f}_n}\left(\frac{\pi\kappa}{A}\right) - \text{Re}\widehat{f}_\theta\left(\frac{\pi\kappa}{A}\right)]^2$ and $E_\lambda E_{f_\theta}[\text{Im}\widetilde{\widehat{f}_n}\left(\frac{\pi\kappa}{A}\right) - \text{Im}\widehat{f}_\theta\left(\frac{\pi\kappa}{A}\right)]^2$, respectively, where $0 < |\kappa| < W$, in the following way:
Take f_θ and $\theta = (\theta_\kappa)_{0<|\kappa|<W}$ from Part III, Section 3, definition (7) and (8), for the the density and parameter in the multivariate van Trees inequality, $p = 2\lfloor W \rfloor$; take $\text{Re}\widehat{f}_\theta\left(\frac{\pi\kappa}{A}\right)$ for $\psi(\theta)$, $s = 1$. Let λ be the prior density defined in Part III, Section 3 definition (9). $B(\theta) \equiv 1$ and let $C(\theta) = (C_{-\lfloor W\rfloor}(\theta_{-\lfloor W\rfloor}), \ldots, C_{-1}(\theta_{-1}), C_1(\theta_1), \ldots, C_{\lfloor W\rfloor}(\theta_{\lfloor W\rfloor}))$ be the unit vector with $C_{|\kappa|}(\theta_{|\kappa|}) = 1$ and all other components 0. Then

$$\begin{aligned}
E_\lambda\left[\text{tr } C(\theta)\left(\frac{\partial\psi}{\partial\theta}\right)^T\right] &= E_\lambda\left[\frac{\partial\text{Re}\widehat{f}_\theta\left(\frac{\pi\kappa}{A}\right)}{\partial\theta}\right] \\
E_\lambda\left[\text{tr } B(\theta)^T C(\theta) I_{f_\theta} C(\theta)^T\right] &= I_{f_\theta}(\theta_{|\kappa|}) \\
I'_\lambda &= E_\lambda\left[\frac{1}{\lambda^2(\theta)}\left(\frac{\partial}{\partial\theta_{|\kappa|}}\lambda(\theta)\right)^2\right] = I_{|\kappa|}
\end{aligned}$$

For $\text{Im}\widehat{f}_\theta\left(\frac{\pi\kappa}{A}\right)$ just take $C(\theta)$ as the unit vector with $C_{-|\kappa|}(\theta_{-|\kappa|}) = 1$ and 0 elsewhere.

B S-PLUS PROGRAM

This is the program that was used for the computation of the CV-optimal kernel with piecewise constant Fourier transform. The intervals, on which $\widehat{K}$ is constant, are slowly growing: length of the k^{th} interval $= a^k b_0$.

Since cross-validation bandwidths are already available in many statistical programs as a standard function, we abstain from providing one here.

```
module(NUOPT)
```

```
n<-500
X<-rnorm(n,0,1)
```

simulate 500 independent standard-normally distributed observations

```
a<-1.1 b0<-1/10
d<-log((n/b0)*(a-1)+1)
d<-floor(d/log(a)-1)
```

divide $[0, n]$ into intervals for the step function; set initial points, end points and step lengths

```
l1<-0:d
l1<-(a^l1)*b0
l2<-cumsum(l1)

l3<-matrix(0,d+2,1)
l3[2:(d+2),1]<-l2

l4<-matrix(n,d+2,1)
l4[1:(d+1),1]<-l2

l5<-matrix(0,d+2,1)
l5<-l4-l3

k<-matrix(1,d+2,1)
```

```
X1<-matrix(X,n,n)
dif<-X1-t(X1)
```

prepare optimization problem for transfer to NUOPT

```
mat1<-sin(kronecker(l4,dif))
mat2<-sin(kronecker(l3,dif))
mat3<-kronecker(k,dif+diag(n))
mat4<-(mat1-mat2)/mat3

mat4<-rowSums(mat4)
mat5<-matrix(mat4,d+2,n,byrow=T)
mat6<-rowSums(mat5)
```

```
stp1<-l5+mat6/n
stp2<-mat6/(stp1*(n-1))
```

set starting point

```
new.model()                                  NUOPT minimizes CV over stepfunctions
nuopt.options(maxitn=500)
optkern<-function(mat6){
    K<-Set(1:(d+2))
    k<-Element(set=K)
    A<-Variable(index=k)
    X<-Parameter(list(1:(d+2),mat6),index=k)
    L<-Parameter(list(1:(d+2),l5),index=k)
    K1<-Parameter(list(1:(d+2),stp2),index=k)
    A[1]<=1
    A[d+2]>=0
    k>1
    A[k]<=A[k-1]
    (k>=1)%&&%(k<=d+2)
    2*Sum(A[k]*A[k]*L[k],k)<=n

    CV1<-Expression(index=k)
    CV1[k]~(A[k]*A[k]*L[k])/n
    CV2<-Expression(index=k)
    CV2[k]~((A[k])*(A[k])/(n*n) - 2*A[k]/(n*(n-1))) * X[k]

    CV<-Objective("minimize")
    CV~Sum(CV1[k],k)+Sum(CV2[k],k)

    A[k]~K1[k]}
sys.optkern<-System(model=optkern,mat6)
sol.optkern<-solve(sys.optkern)

lsg<-as.list(current(sys.optkern,A))
K0<-lsg$values
```

compute the estimate $\widetilde{f}_{K_0}$ (for example) at point $x = 1$

```
x<-1
X2<-matrix(X,1,n,byrow=T)
X2<-X2-x

mat7<-sin(kronecker(l4,X2))
mat8<-sin(kronecker(l3,X2))
mat9<-kronecker(k,X2)
mat10<-(mat7-mat8)/mat9

mat10<-rowSums(mat10)

fK0tilde<-sum(mat10*K0)/(n*pi)
```

estimate for f at point $x = 1$

References

[1] Arcones, M.A. and Giné, E. (1993). Limit theorems for U-processes. *A. Probab.* **21** No.3, 1494-1542.

[2] Arcones, M.A. and Giné, E. (1995). On the law of iterated logarithm for canonical U-statistics and processes. *Stochastic Processes Appl.* **58** No.2, 217-245.

[3] Bentkus, R. and Rudzkis, R. (1980). On exponential estimates of the distribution of random variables. *Litov. Mat. Sb.* **20** No.1, 15-30. (Russian)

[4] Borell, C. (1979). On the integrability of Banach space valued Walsh polynomials. Séminaire de Probabilité XIII, Univ. Strasbourg 1977/78 *Lect. Notes Math.* **721**, 1-3.

[5] Borovkov, A.A. and Sakhanenko, A.U. (1980). On estimates of the expected quadratic risk. *Probab. Math. Statist.* **1**, 185-195.

[6] Borovkov, A.A. (1984). *Mathematical statistics. Estimation of parameters. Test of hypothesis.* Nauka. Moscow. (Russian)

[7] Bowman, A.W. (1984). An alternative method for cross validation for the smoothing of density estimators. *Biometrika* **71** 353-360.

[8] Chandrasekharan, K. (1989). *Classical Fourier transforms* Springer-Verlag. Berlin.

[9] Davis, K.B. (1977). Mean integrated square error properties of density estimates. *Ann. Statist.* **5**, 530-535.

[10] de la Peña, V.H. (1992). Decoupling and Khintchine's inequality for U-statistics. *Ann. Probab.* **20** No.4, 1877-1892.

[11] de la Peña, V.H. and Montgomery-Smith, S.J. (1995). Decoupling inequalities for the tail probabilities of multivariate U-statistics. *Ann. Probab.* **23** No.2, 806-816.

[12] Efroimovich, S.Yu. and Pinsker, M.S. (1983). Estimation of square-integrable probability density of a random variable. *Probl. Inf. Transm.* **18**, 175-189.

[13] Epanechnikov, V.A. (1969). Nonparametric estimation of a multidimensional probability density. *Teor. Veroyatn. Primen.* **14**, 156-162. (Russian)

[14] Gasser, T., Müller, H.-G. and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *J. R. Stat. Soc., Serie B* **47**, 238-252.

[15] Gill, R.D. and Levit, B.Y. (1995). Applications of the van Trees inequality: A Bayesian Cramér-Rao bound. *Bernoulli* **1** No.1-2, 59-79.

[16] Golubev, G.K. (1987). Adaptive asymptotically minimax estimators of smooth signals. *Probl. Inf. Transm.* **23**, 47-55.

[17] Golubev, G.K. (1991). LAN in problems of nonparametric estimation of functions and lower bounds for quadratic risk. *Theory Probab. Appl.* **36** No.1, 152-157.

[18] Golubev, G.K. (1992). Nonparametric estimation of smooth probability densities in L_2. *Probl. Inf. Transm.* **28** No.1, 44-54.

[19] Golubev, G.K. and Levit, B.Y. (1996). On the second order minimax estimation of distribution functions. *Math. Methods Stat.* **5** No.1, 1-31.

[20] Höffding, W. (1963). Probability inequalities for sums of bounded random variables. *JASA* **58**, 13-30.

[21] Hall, P. (1983). Large sample optimality of least squares cross-validation in density estimation. *Ann. Statist.* **11**, 1156-1174.

[22] Hall, P. (1984). Asymptotic properties of integrated square error and cross-validation for kernel estimation of a regression function. *Z. Wahrscheinlichkeitstheorie Verw. Geb.* **67**, 175-196.

[23] Hall, P. and Marron, J.S. (1987). Extent to which least squares cross-validation minimizes integrated square error in nonparametric density estimation. *Probab. Theory Relat. Fields* **74**, 567-581.

[24] Hall, P. and Marron, J.S. (1987). On the amount of noise inherent in bandwidth selection for a kernel density estimator. *Ann. Statist.* **15**, 163-181.

[25] Hall, P. (1993). On plup-in rules for local smoothing of density estimators. *Ann. Statist.* **21**, 694-710.

[26] Hansen, B.E. (2003). Exact mean integrated squared error of higher order kernel estimators. Working paper (www.ssc.wisc.edu/~bhansen).

[27] Johnson, N.L. and Kotz, S. (1970). *Distributions in statistics.* Vol. 1. John Wiley & Sons. New York.

[28] Kneip, A. (1994). Ordered linear smoothers. *Ann. Statist.* **22** No.2, 835-866.

[29] Lütkepohl, H. (1996). *Handbook of matrices.* John Wiley & Sons. Chichester.

[30] Marron, J.S. and Park, B.U. (1990). Comparison of data-driven bandwidth selectors. *JASA* **85** Vol. 409, Theory and Methods, 66-72.

[31] Nolan, D. and Pollard, D. (1987). U-processes: rates of convergence. *Ann. Statist.* **15**, 780-799.

[32] Nolan, D. and Pollard, D. (1988). Functional limit theorems for U-processes. *Ann. Probab.* **16** No.3, 1291-1298.

[33] Petrov, W.W. (1987). *Limit theorems for sums of independent random variables.* Nauka. Moscow. (Russian)

[34] Pollard, D. (1984). *Convergence of stochastic processes.* Springer Series in Statistics. New York.

[35] Rigollet, P. (2004). Adaptive density estimation using Stein's blockwise method. Preprint PMA-913 (available at www.proba.jussieu.fr)

[36] Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scand. J. Stat., Theory Appl.* **9**, 65-78.

[37] Samko, G. (1987). *Fractional integrals and derivatives: theory and applications.* Gordon and Breach Science Publishers. Yverdon.

[38] Saulis, L. and Statulevičius, V. (1991). *Limit theorems for large deviations.* Kluwer Academic Publishers, Dordrecht.

[39] Schipper, M. (1996). Optimal rates and constants in L_2-minimax estimation of probability density functions. *Math. Methods Stat.* **5** No.3, 253-274.

[40] Serfling, R.J. (1980). *Approximation theorems of mathematical statistics.* John Wiley & Sons. New York.

[41] Shorack, G.R. and Wellner, A.J. (1986). *Empirical Processes with applications to statistics.* John Wiley & Sons. New York.

[42] Stone, C.J. (1984). An asymptotically window selection rule for kernel density estimates. *Ann. Statist.* **12**, 1285-1297.

[43] van Trees, H.L. (1968). *Detection, estimation and modulation theory. Part I* John Wiley & Sons. New York - London - Sydney.

[44] Zinn, J. (1985). Comparison of martingal difference sequences. Probability in Banach spaces V, Proc. 5th Int. Conf., Medford/Mass. 1984 *Lect. Notes Math.* **1153**, 453-457.